



Transport Canada
Policy and
Coordination

Research and
Development

Transports Canada
Politiques et
coordination

Recherche et
développement

TP 9983

Transportation
Development Centre

Centre de développement
des transports

ELECTRONIC SYSTEMS IN TRANSPORTATION

SYSTÈMES ÉLECTRONIQUES EN TRANSPORT

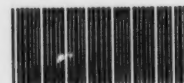
PROCEEDINGS

COMPTE RENDU

CA1
DTC800-
89
E47

427342

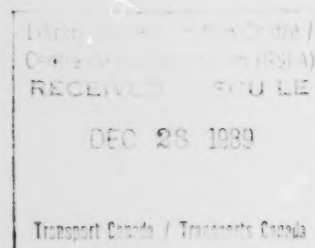
41.D



Edited by/Édité par
Micha Avni

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRÈS CANADIEN EN GÉNIE ÉLECTRIQUE ET INFORMATIQUE

Le Centre Sheraton, Montréal, Canada



427342

The contents of the papers included in the proceedings reflect the views of the authors and not necessarily the official views or opinions of the Transportation Development Centre of Transport Canada.

The papers are published in the language in which they were presented at the conference.

Le contenu de ces documents n'engage que leurs auteurs et ne correspond pas nécessairement aux opinions ou aux vues du Centre de développement des transports, Transports Canada.

Les communications sont publiées dans la langue où elles ont été présentées.

Single reprints of any part of this publication may be made for the purpose of private study, research, criticism or review provided exact reference thereto is quoted. Such reprints may not be offered for sale.

Printed in Montréal.

Des réimpressions de toute partie de cette publication peuvent être faites pour étude privée, recherche, critique ou revue, pourvu que la source exacte en soit citée. Ces réimpressions ne peuvent être mises en vente.

Imprimé à Montréal.

1. Transport Canada Publication No. TP 9983		2. Project No. 6902		3. Recipient's Catalogue No.	
4. Title and Subtitle Proceedings — Electronic Systems in Transportation				5. Publication Date September 1989	
				6. Performing Organization Document No. —	
7. Author(s) M. Avni (Editor)				8. Transport Canada File No. D 1450-121-6	
9. Performing Organization Name and Address Communications Gillies Inc. 410 Walnut Ave. St. Lambert, Quebec J4P 2T6				10. DSS File No. —	
				11. DSS or Transport Canada Contract No. —	
12. Sponsoring Agency Name and Address Transportation Development Centre (TDC) Guy Favreau Complex 200 René Lévesque Blvd. West West Tower, Suite 601 Montreal, Quebec H2Z 1X4				13. Type of Publication and Period Covered Proceedings, 1989	
				14. Project Officer Micha Avni	
15. Supplementary Notes (Funding programs, titles of related publications, etc.)					
16. Abstract This report presents the papers relating to Electronic Systems in Transportation given at the Canadian Conference on Electrical and Computer Engineering held at Le Centre Sheraton, Montreal, Quebec, September 17-20, 1989. The conference was sponsored by the Canadian Society for Electrical and Computer Engineering, a constituent of the Engineering Institute of Canada, in cooperation with the IEEE Montreal Section, the IEEE Computer Society and the IEE (UK). The co-chairmen of the conference were Paul-E. Allard of the Royal Military College, Kingston, and Micha Avni of the Transportation Development Centre of Transport Canada.					
17. Key Words Proceedings			18. Distribution Statement Limited number of copies available from the Transportation Development Centre		
19. Security Classification (of this publication) Unclassified	20. Security Classification (of this page) Unclassified	21. Declassification (date) —	22. No. of Pages 151	23. Price —	

1. N° de la publication de Transports Canada TP 9983		2. N° de l'étude 6902		3. N° de catalogue du destinataire	
4. Titre et sous-titre Compte rendu — Systèmes électroniques en transport				5. Date de la publication Septembre 1989	
				6. N° du document de l'organisme exécutant —	
7. Auteur(s) M. Avni (éditeur)				8. N° de dossier — Transports Canada D 1450-121-6	
9. Nom et adresse de l'organisme exécutant Communications Gillies Inc. 410 avenue Walnut St-Lambert (Québec) J4P 2T6				10. N° de dossier — ASC —	
				11. N° de contrat — ASC ou Transports Canada —	
12. Nom et adresse de l'organisme parrain Centre de développement des transports (CDT) Complexe Guy-Favreau 200 ouest, boul. René-Lévesque Tour ouest, suite 601 Montréal (Québec) H2Z 1X4				13. Genre de publication et période visée Compte rendu, 1989	
				14. Agent de projet Micha Avni	
15. Remarques additionnelles (Programmes de financement, titres de publications connexes, etc.)					
16. Résumé Ce rapport reproduit les mémoires relatifs aux systèmes électroniques en transport présentés au Congrès canadien en génie électrique et informatique qui s'est tenu au Centre Sheraton, Montréal (Québec) du 17 au 20 septembre 1989. Cette conférence était commanditée par la Société canadienne de génie électrique et informatique, constituante de l'Institut canadien des ingénieurs, en collaboration avec IEEE Montréal, filiale de IEEE Computer Society et IEE (UK). La coprésidence était assurée par Paul-E. Allard du Collège militaire royal de Kingston et Micha Avni du Centre de développement des transports, Transports Canada.					
17. Mots clés Compte rendu			18. Diffusion Le Centre de développement des transports dispose d'un nombre limité d'exemplaires.		
19. Classification de sécurité (de cette publication) Non classifiée	20. Classification de sécurité (de cette page) Non classifiée	21. Déclassification (date) —	22. Nombre de pages 151	23. Prix —	



Minister
Transport Canada

Ministre
Transports Canada



Advanced electronic systems have a growing number of applications in the aviation, marine and railway industries. These systems find their way into the vehicles of all modes, into the transportation infrastructure, and into the management and administration of the transportation sector. They improve the safety and efficiency of the transportation network in Canada and provide Canadians with more productive and economical transportation services.

The Transportation Development Centre (TDC), Transport Canada's research and development organization, is increasing its R&D efforts in the area of electronic and computer systems. Several of the papers included in these proceedings present the results of TDC's collaboration with industry and university researchers in support of Transport Canada and the transportation community.

We look forward to continued progress and innovation in this domain.

Les applications des systèmes électroniques de pointe ne cessent de se multiplier dans les industries de l'aviation, de la marine et des chemins de fer. Aujourd'hui, on trouve ces systèmes dans tous les types de véhicule, dans l'infrastructure des transports et dans les centres de gestion et d'administration du secteur des transports. Ils aident à améliorer la sécurité et l'efficacité des réseaux de transport au Canada et à fournir des services d'une façon plus productive et économique aux Canadiens.

Le Centre de développement des transports (CDT), l'organisme de recherche et de développement de Transports Canada, intensifie ses efforts de R&D dans le domaine des systèmes électroniques et informatiques. Dans ces comptes rendus, plusieurs exposés sont le fruit de la collaboration du CDT avec les chercheurs de l'industrie et du milieu universitaire dans le cadre de projets profitant à Transports Canada et au secteur des transports.

Fort de cette collaboration, ce secteur continuera d'évoluer et d'innover.

Benoît Bouchard
Minister of Transport
Ministre des Transports

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING/ CONGRÈS CANADIEN EN GÉNIE ÉLECTRIQUE ET INFORMATIQUE

EXECUTIVE COMMITTEE/COMITÉ EXÉCUTIF DU CONGRÈS

Conference Co-Chairmen/Co-présidents du congrès
Paul-E. Allard
Micha Avni

Technical Program Chairman/Président du programme technique
Vijay K. Bhargava

Finance Chairman/Trésorier
Joseph Fox

General Secretary/Secrétaire général
Pierre Allard

Registration Chairman/Président des inscriptions
Jean Conan

Publicity Chairman/Président de la publicité
George Ekins

Publications Chairman/Président des publications
Tho Le-Ngoc

Sponsored by the Canadian Society for Electrical and Computer Engineering,
a constituent of the Engineering Institute of Canada

Commandité par la Société canadienne de génie électrique et informatique,
constituante de l'Institut canadien des ingénieurs

In cooperation with/En collaboration avec

IEEE Montréal Section de Montréal

IEEE Computer Society

IEE (UK)

Financially Supported by/financé par

Transportation Development Centre/Centre de développement des transports, Transport Canada/Transports
Canada • Canadian Workplace Automation Research Centre/Centre canadien de recherche sur
l'informatisation du travail, Communications Canada • Ministère des communications, Gouvernement du
Québec • NSERC/CRSNG • Bell Canada • Microtel Pacific Research Limited • Bell-Northern
Research/Recherche Bell-Northern • Canadian Marconi • Raytheon Canada • Alberta Telecommunications
Research Centre • Virtual Prototypes Inc. • Hydro-Québec • Centre de recherche informatique de Montréal

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING/ CONGRÈS CANADIEN EN GÉNIE ÉLECTRIQUE ET INFORMATIQUE

WELCOME FROM THE CONFERENCE CO-CHAIRMEN

On behalf of the organizing committee, we are pleased to welcome you to Montreal for the second Canadian Conference on Electrical and Computer Engineering (1989). The theme of this year's conference "*Canada in the information age*" is in recognition of the importance of the information technologies to Canada's future and to that of the world community. The large number of excellent papers that will be presented demonstrate the liveliness of Canadian researchers working in the many diverse areas of Electrical and Computer Engineering. We are confident that the conference offers an outstanding opportunity for a fruitful exchange of ideas and renewal of contacts amongst researchers from Canada and overseas.

Many individuals have contributed to this event. Special thanks are due to our colleagues on the organizing committee who gave their time and energy willingly and enthusiastically; we are most grateful for their efforts. The cooperation received from the corporate and public sector supporters is also gratefully acknowledged. Special thanks are due to Dr. Vijay Bhargava who demonstrated at the first CCECE (Vancouver, November 1988) that such a forum was needed and who again provided invaluable assistance in the preparation of this second conference.

Finally, we thank you for your participation and wish you a pleasant and profitable conference. We hope to meet you all at the next CCECE to be held in the Ottawa Congress Centre, September 11-14, 1990.

Paul-E. Allard and Micha Avni
Conference Co-Chairmen

MOT DE BIENVENUE DES CO-PRÉSIDENTS DU CONGRÈS

Au nom du comité organisateur, nous sommes heureux de vous accueillir à Montréal, pour le deuxième Congrès canadien en génie électrique et informatique (1989). Le thème du congrès de cette année "*Canada à l'âge de l'information*" a été choisi en reconnaissant l'importance des technologies de l'information pour le futur du Canada et pour celui du monde entier. Le grand nombre de communications qui seront présentées, démontre l'excellence des chercheurs canadiens travaillant dans les différentes sphères du génie électrique et informatique. Nous sommes persuadés que le congrès offre une occasion sans pareille pour un partage fructueux d'idées et un renouvellement de contacts au profit des chercheurs du Canada et d'outre-mer.

Plusieurs personnes ont contribué à cet événement. Des remerciements bien spéciaux s'adressent à nos collègues du comité organisateur qui ont donné temps et énergie avec enthousiasme; nous leur sommes très reconnaissants pour leurs efforts. La coopération des services publics et corporatifs mérite également l'expression de notre gratitude. Nous remercions tout particulièrement Vijay Bhargava qui a démontré lors du premier CCGEI (Vancouver, novembre 1988), la nécessité d'un tel forum et qui nous a assisté de façon inestimable dans la préparation de ce second congrès.

Nous vous remercions de votre participation en vous souhaitant un congrès plaisant et fructueux. Nous espérons vous revoir au prochain CCGEI qui se tiendra du 11 au 14 septembre 1990 au Centre des congrès d'Ottawa.

Paul-E. Allard et Micha Avni
Co-présidents du congrès

ix TABLE OF CONTENTS/TABLE DES MATIÈRES

TRANSPORTATION DEVELOPMENT CENTRE/ CENTRE DE DÉVELOPPEMENT DES TRANSPORTS

TDC: Innovation in Transport N.E. Rudback, TDC/CDT, Montreal, PQ	2
Marine Applications of Electronics and Computers J. Reid, TDC/CDT, Montreal, PQ	*
Aviation Applications of Electronics and Computers M. Avni, TDC/CDT, Montreal, PQ	*
Railway Transportation Applications of Electronics and Computers W.S.C. McLaren, TDC/CDT, Montreal, PQ	*
Highway Transportation Applications of Electronics and Computers B. Marshall, TDC/CDT, Montreal, PQ	*

* oral presentation only

MLS/DME

Guidance Angle Measurement of an MLS Antenna S. Radhakant, B.C. Blevis, Canadian Astronautics Limited, Ottawa, ON R. Ulan, M.Z. El-Gamal, Transport Canada, Ottawa, ON	6
Image-Supported Navigation for Testing Instrument Landing Systems U. Brokof, F. Knabe, DLR Braunschweig, Braunschweig, Federal Republic of Germany	10
Comprehensive MLS Testing for Category I, II & III Operation A.E. Smith, MSI Services Inc., Washington, DC, U.S.A.	14
The Effect of Error Tolerances on MLS Array Antenna M.Z. El-Gamal, C. Bowman, Transport Canada, Ottawa, ON	18
MLS System Error Model Identification and Synthesis R.J. Kelly, Allied-Signal Aerospace, Baltimore, MD, U.S.A.	22
Factory Acceptance Test Range R. Janes, M. MacNeil, Micronav Limited, Sydney, NS	29
Design and Analysis Techniques for MLS Safety Critical Software S.H. Houston, M. Vickers, Lisan Corporation, Washington, DC, U.S.A.	32
DME Growth Elements and Their Use with MLS A. Becker, Institut für Flugführung, Braunschweig, Federal Republic of Germany	36

x

ELECTRONIC SYSTEMS IN TRANSPORTATION SYSTÈMES ÉLECTRONIQUES EN TRANSPORT



MLS (cont'd)/DME (suite)

DME/P Transponder Reply Delay Accuracy - Factory Verification and Calibration E. Nisson, E-Systems Inc., Salt Lake City, UT, U.S.A.	40
Testing MLS DME/P Transponders and Interrogators P.E. Manning, Department of Transport, Cambridge, MA, U.S.A. E.J. Pesciotta, J.W. Ruggiero, UNISYS Corporation, Cambridge, MA, U.S.A.	45
Validation Testing of DME/P Ground Transponders and Airborne Interrogators J.F. Michaels, Republic Electronics Co., Hauppauge, NY, U.S.A.	49
A DME Transponder System Approach for Convertibility, Monitoring and Diagnostics T.A. Hicks, Leigh Instruments Limited, Kanata, ON	53
Testing of CMA-2000 Microwave Landing System (MLS) Airborne Receiver L. Labrèche, A.J. Murfin, Canadian Marconi Company, Montreal, PQ	57
MLS Testing Using Built-in Test R. Janes, D. Babineau, Micronav Limited, Sydney, NS	60
A Strategy for Exploiting the Full Potential of MLS based Terminal Procedures in Canada K.R. Singh, Aviation Planning Services Ltd., Montreal, PQ	63
Near Filed Receiving Properties of Printed Loop Antennas Z.A. Delecki, M. Barakat, H. Kroker, National Research Council Canada, Winnipeg, MB L. Shafai, M. Pawlak, University of Manitoba, Winnipeg, MB L. Soiferman, Wardrop Engineering Inc., Winnipeg, MB	69

ELECTRONIC SYSTEMS IN SEARCH AND RESCUE/ L'ÉLECTRONIQUE DANS LES OPÉRATIONS DE RECHERCHE ET DE SAUVETAGE

A Computer-Integrated Radar System for Research Use D. Hamburger, C. Krasnor, S. Haykin, B. Currie, T. Nohara, McMaster University, Hamilton, ON	74
Radar Vision S. Haykin, McMaster University, Hamilton, ON	77

ELECTRONIC SYSTEMS IN TRANSPORTATION SYSTÈMES ÉLECTRONIQUES EN TRANSPORT



ELECTRONIC SYSTEMS IN SEARCH AND RESCUE (cont'd)/ L'ÉLECTRONIQUE DANS LES OPÉRATIONS DE RECHERCHE ET DE SAUVETAGE (suite)

CANSARP - A Computerized Search and Rescue Planning Program for the Canadian Coast Guard	81
B.C. Berglund, D.O. Hodgins, Seaconsult Marine Research Ltd., Vancouver, BC	
Waveform Design for a High Frequency Ocean Surveillance Radar	85
R.H. Khan, Memorial University of Newfoundland, St. John's, NF	
D.K. Mitchell, Northern Radar Systems Limited, St. John's, NF	
The Potential Use of Radar Technology in Ship Based Search and Rescue Operations	89
B.R. Dawe, NORDCO Limited, St. John's, NF	
M. Kelly, Canadian Coast Guard, Ottawa, ON	
C.A. Versailles, Transportation Development Centre, Montreal, PQ	

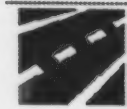
ADVANCED SYSTEMS FOR AIR SPACE MANAGEMENT/ SYSTÈMES AVANCÉS DE GESTION DE L'ESPACE AÉRIEN

A Taxi and Ramp Management and Control System (TARMAC)	94
U. Völckers, German Aerospace Research Establishment, Braunschweig, Federal Republic of Germany	
Automated Problem Resolution Prototype in Automated En Route Air Traffic Control	99
C.V. Fong, The MITRE Corporation, McLean, Va, U.S.A.	
A Unique Approach to Aircraft Conflict Resolution Using Artificial Intelligence Techniques	103
I.G. Smith, C.A. Crabb, G.R. Ram, M.R. Attisha, CompEngServ Ltd., Ottawa, ON	

OTHER PAPERS OF INTEREST/ AUTRES PRÉSENTATIONS D'INTÉRÊT

Sliding Mode Motion Control of an Autonomous Vehicle	108
R.M. DeSantis, R. Hurteau, École Polytechnique de Montréal, Montréal, PQ	

ELECTRONIC SYSTEMS IN TRANSPORTATION SYSTÈMES ÉLECTRONIQUES EN TRANSPORT



OTHER PAPERS OF INTEREST (cont'd)/ AUTRES PRÉSENTATIONS D'INTÉRÊT (suite)

Coherent Detection of Interleaved Trellis Encoded CPFSK for Shadowed Mobile Satellite Channels	112
R.W. Kerr, P.J. McLane, Queen's University, Kingston, ON	
Wide Area Vehicular Communications	118
A.I. Spolsky, Microtel Pacific Research Limited, Burnaby, BC	
M.O. Ali, South West Bell, Kansas City, MI, U.S.A.	
V.C.M. Leung, University of British Columbia, Vancouver, BC	
A Hard Real-Time Operating System for Airborne Scientific Experiments	123
J.E. Jordan, National Research Council Canada, Ottawa, ON	
Feasibility of Using Segmentation Techniques for the Enhancement of Marine Radar Images as Encountered in Ice Infested Waters	127
M. Huculak, A. Nilakantan, Cimmetry Systems, St. Laurent, PQ	
C. Gautier, Canarctic Shipping, Ottawa, ON	
M. Audette, Transportation Development Centre, Montreal, PQ	
Logis-Cell Array Implementation of a (31,k) Reed-Solomon Codec and it's Applications	131
O. Dravnieks, W.D. Little, V.K. Bhargava, University of Victoria, Victoria, BC	
Automated Baggage Threat Detection	135
T.E. Feather, A. Lee-Kwen, L. Guan, S. Berkowitz, Array Systems Computing Inc., Downsview, ON	
M. Audette, Transportation Development Centre, Montreal, PQ	
A Graphic Integrated Software Package for Road Management	139
A. Bedetti, SOFOS s.r.l., Imola, Italy	
R. Daverio, Technisoft, Varese, Italy	
B. DiStefano, E. DiStefano, Nuptek Systems Ltd., Toronto, ON	
Effective Interface to Expert Systems in Transportation	143
A. Griffith, T. Gomi, K. Mayer, U. Oppacher, Applied AI Systems Inc., Kanata, ON	
L. Sabounghi, Transportation Development Centre, Montreal, PQ	
An Expert System in Railway Information System	147
A. Pozniak, A. Luk, CN Rail, Montreal, PQ	
AUTHOR INDEX AUTEUR	151

**TRANSPORTATION DEVELOPMENT CENTRE/
CENTRE DE DÉVELOPPEMENT DES TRANSPORTS**

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

TDC: INNOVATION IN TRANSPORT

N.E. Rudback
Executive Director

Transportation Development Centre
Guy Favreau Complex
200 René-Lévesque West, West Tower, Suite 601
Montreal, Quebec H2Z 1X4

Abstract

The Transportation Development Centre (TDC) manages Transport Canada's core R&D program. Federal government funding of this program is currently at a level of \$14.5 M annually. The program addresses all modes of transport with the R&D work being contracted out to the private sector. The paper outlines the scope of the R&D program and the methods used in defining, implementing and managing the R&D activities.

Introduction

The Transportation Development Centre, formed in 1970, is Transport Canada's centralized research and development (R&D) organization. TDC formulates, manages, or reports on an average of 300 R&D projects a year. These projects are contracted to a variety of organizations, most of them in the private sector, and cover all transportation modes, systems, and technologies. TDC serves as Transport Canada's centre of expertise on transportation technology and innovation.

In February 1988, Transport Canada announced a policy of sustained commitment to the R&D required to meet the department's and the transportation sector's needs for technological innovation through sustained and stable funding for a Core R&D Program. The administration of the program has been placed in the hands of the Assistant Deputy Minister, Policy and Coordination, aided by a new top-level R&D Council that will provide advice on objectives and priorities.

The chairman of the R&D Council is the Director General, Research and Development, Policy and Coordination Group. There are twelve members representing Transport Canada's operating groups and TDC. Among other things, the new Council enables long-range priorities to be established and research to be carried out on a sustained basis. R&D can now go further into the future since it is not restricted by the ups and downs of the annual budget cycle.

TDC reports to the Director General, Research and Development, and works in accordance with directives and guidelines of the Core R&D Program. These directives establish R&D priorities and the requirements in support of Transport Canada's activities and of federal science and technology policy. They also reflect requirements as expressed through transportation R&D advisory boards.

TDC Role and Mission

The role of TDC can be described as "The application of science and technology toward a safer, more efficient and effective Canadian transportation system".

In carrying out this broad role, the R&D activities are divided into three categories corresponding to the three mission requirements which are addressed by the Core R&D Program.

- **Long-Term Applications R&D** to provide other components of Transport Canada with specific data, research, development and demonstration projects to assist them in their operational, regulatory and policy planning functions. This type of activity addresses areas where the department has a direct operating responsibility and includes support of capital investment programs. Examples of this activity are improved air traffic control equipment and systems, improved ice breakers and ice breaking technology, and increased safety in the regulation of dangerous goods transport.

- **Sector-Directed R&D** supports innovations in the transportation sector at large that further Transport Canada's objectives and priorities. The R&D programs are determined by policy directives for safety, productivity, regional development, and environmental protection. Long term safety and productivity issues relating to air, marine and rail modes are some of the concerns sector-directed R&D may address. These kinds of programs involve cooperation with the provinces, industry and other federal departments and agencies.

- **Exploratory R&D** aims to produce knowledge through systematic assessment of new developments. It does this by identifying technological innovations with the potential to improve elements of transportation systems and encouraging or supporting the development of appropriate technologies. Methods include: technology scanning, forecasting and assessment; experimental development and testing; and consultative workshops and symposia. Investigations frequently are high risk, require long-term monitoring, and involve a wide range of users, R&D performing organizations and interest groups.

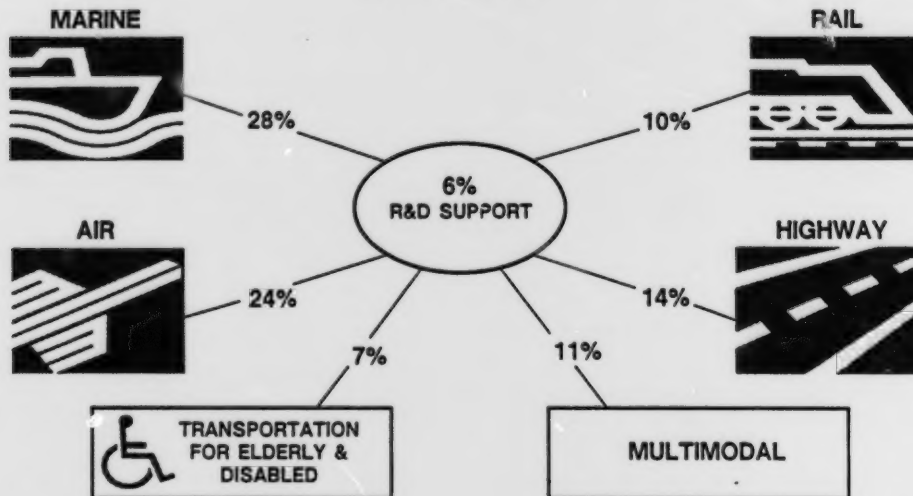
The research is carried out by TDC in close cooperation with its colleagues in the operating groups of Transport Canada: Airports authority, Aviation, Marine and Surface. They contribute directly to this effort by helping to define the R&D objectives, reviewing the technical work, and by participating on project steering committees. This valuable assistance is acknowledged as a key element in the success of the Core R&D Program. The allocation of funding by R&D category is illustrated in Figure 2.

Funding

Transport Canada has made a commitment to a sustained long term core research and development program. The department dedicates 1.25% of its capital budget or \$9 M (whichever is greater) to the core R&D effort. This core budget is augmented through cost sharing with other Transport Canada components, other federal departments, industry or the provinces.

Major contributors to R&D projects described in the TDC Project Directory are the Department of Energy, Mines and Resources (EMR) through its Energy R&D Program; the Cabinet Committee for Economic and

Figure 1
TDC's R&D Program, 1989-90: Program Structure
(\$14.5 million*)



* Excludes Canada-Quebec ERDA projects and industry and provincial contributions

Regional Development through its Transportation for Disabled Persons (TDPP); the federal government and the province of Quebec through an Economic and Regional Development Agreement (ERDA) sub-agreement on transportation development. Other federal government funding sources add \$5.5 M to the core R&D budget, bringing the total federal support to \$14.5 M.

Cost sharing with industry and the provinces through direct financial payments or contributions in kind are estimated at about \$5.5 M, resulting in an annual R&D program valued at \$20 M. Cost sharing with industry is the norm when there is a reasonable potential of future commercial benefits for the company carrying out the work. It also helps ensure that the company carrying out the work has a vested interest in the outcome, providing incentive for a high quality product. Safety R&D and environmental protection activities of prime interest to Transport Canada generally receive 100% federal government funding.

Human Resources

Forty full time staff employed at TDC in Montreal are responsible for the program delivery function. About twenty-seven of the staff are technical professionals, principally TDC project officers. The TDC project officers are highly skilled and experienced technical specialists with a broad understanding of transportation problems. They are involved in all stages of the project cycle from project definition, through identification of funding sources, seeking of support and approval, request for contracts, evaluation of proposals, technical monitoring of the work, and interpretation of the research results. The success of a project is often dependent upon the dedication and skill of the project officer in directing the R&D work, overcoming technical difficulties and in convincing the operating personnel of the advantages in adapting new methods.

The technical expertise at TDC is supplemented by a Visiting Experts Program which provides recognized

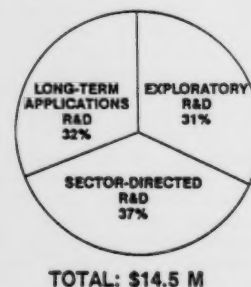
technical experts the opportunity to carry out two to six months research assignments at TDC. Experts appointed under the program interact with TDC staff and others in Transport Canada on major transport problems, and their knowledge and experience provides valuable guidance to TDC staff.

Policy direction, overall research planning and coordination support for Transport Canada's R&D program is provided by staff of 15 persons located in Ottawa. This support allows TDC to direct most of its attention toward the project planning and implementation.

R&D Program

The R&D program addresses all modes of transport: marine, air and surface. Special attention is also given to the transport of elderly and disabled persons as well as the multimodal aspects of intercity transport. The development of the transportation technology base, common to all modes of transport, is included in the multimodal R&D activities. Research allocation among the transport modes is illustrated in Figure 1.

Figure 2
Funding Breakdown by R&D category



The R&D program is carried out on a project basis, with each project having its own specific objective, budget and schedule. Approximately 300 projects are underway at any given time. A description of some of the major activities in the marine, air, highway and rail modes will be provided by other speakers on the conference program. A number of current projects are also described in the TDC Project Directory (Reference 1). The overall direction of the R&D program is detailed in the TDC Operational Plan (Reference 2). The Operational Plan provides an outline of planned R&D scheduled for implementation in the current fiscal year.

Contracting out of R&D

The R&D activities of the Transportation Development Centre are contracted out to the private sector; universities, consultants and manufacturing firms. The contracting out policy permits TDC to manage a large R&D program with a small core of highly skilled professionals.

In implementing its projects, TDC utilizes the best available technical expertise, wherever it exists across Canada. Where suitable expertise is not available in an area of new technology, TDC participates in the establishment of this expertise through appropriate research contracts. Contracting out avoids the problem of technology transfer from a government laboratory to the private sector. The technical capability, from the outset, is in the private domain where it can be fully exploited for transportation or other applications.

TDC collaborates closely with other components of Transport Canada in technological developments to meet the department's operational needs for state-of-the-art equipment and systems. However, the procurement of the equipment to be used by Transport Canada is done by the operating groups alone, through an open competitive bidding process. This requires that companies working with TDC develop world class products, meeting the highest performance and safety standards at a competitive price. The high quality of the results produced is indicated by a number of Canada Awards for Excellence won by TDC contractors, as well as the high performance standards of equipment delivered to Transport Canada.

Collaboration with Others

The challenges of the transportation sector are too great to be fully addressed by one organization such as TDC operating alone. This is recognized by TDC in carrying out its mandate, and strong emphasis is placed on collaboration with others, both in the public and private sector as well as through international science and technology agreements.

First and foremost, TDC works with the Transport Canada operating groups in Marine (Canadian Coast Guard), Aviation, Airports and Surface. Industry input on the R&D priorities of the private sector is obtained through R&D advisory boards which provide advice to the department through the responsible Assistant Deputy Ministers.

Reliance is placed on the National Research Council (NRC) for basic and applied research. Transport Canada has a formal agreement with NRC which ensures that the development work carried out in the NRC laboratories is well coordinated with work carried out at TDC and elsewhere in Transport Canada. Likewise, there is close coordination with industry support

programs in the transport sector which are funded by the Department of Industry, Science and Technology (DIST). On a number of projects, such as the microwave landing system, NRC, TDC and DIST have collaborated to establish the basic technical expertise in Canada to bring the innovation through from applied research to operational deployment.

It is in the interest of Transport Canada to avoid duplication of research effort and, where appropriate, to adapt technological developments elsewhere to meet our own requirements rather than "reinventing the wheel". Formal collaboration takes place under agreements with Ontario, Quebec and other provinces as well as through international science and technology agreements. Such agreements exist with many countries, including the U.S.A., U.K., France, Germany and Japan. The most active of these is the "Memorandum of Understanding" between Transport Canada and the U.S. Department of Transportation.

Results: Innovation in Transport

TDC places strong emphasis on ensuring that the research results are applied to introduce operational improvements into the transportation systems. Until research reaches the stage of operational deployment, it is of limited value. TDC points with pride to its many contributions it has made to improve transportation technology. The program has directly contributed to innovations in all modes of transport.

A few examples of technical innovation in which TDC played a major role include the establishment in Canada of world leadership in the design of efficient icebreaking ships; the world's first ice navigation system for finding the optimum route through Arctic ice under conditions of 24 hour darkness; development of a microwave landing system (MLS) to meet aviation requirements of the 1990's; advanced technology articulated intercity buses now operating between Montreal and Quebec City; electronic train control and cabooseless train operation. Some highlights of TDC research results can be found in the Annual Review and in Selected R&D Achievements (Reference 3 and 4).

These few examples illustrate the positive benefits which can flow from sustained R&D, and illustrate the innovative capabilities of Canadians in resolving the national problem of safe and efficient transport. Innovation in transport is a key to economic success and prosperity for this immense land of relatively sparse population. TDC, through the Core R&D Program, is continuing to encourage transportation innovation and excellence.

References

1. TDC Project Directory, 10th Edition, January 1989, Transportation Development Centre, Policy & Coordination Group, Transport Canada (TP 1936)
2. TDC Operational Plan 1989-90, August 1989 (TP 5474)
3. TDC Annual Review 1988-89, September 1989 (TP 3230)
4. Selected R&D Achievements, October 1987 (TP 8786)

MLS/DME

GUIDANCE ANGLE MEASUREMENT OF AN MLS ANTENNA

S. Radhakant*, B. C. Blevis*, R. Ulan**, and M. Z. El-Gamal**

* Canadian Astronautics Limited, 1050 Morrison Drive, Ottawa,
Ontario, Canada K2H 8K7

** Microwave Landing System Project, Transport Canada,
110 O'Connor Street, Ottawa, Ontario, Canada K1A 0N8

ABSTRACT

With the impending installation of Microwave Landing Systems (MLS) at major airports of the world, it is essential to devise means for testing the accuracy of MLS antennas and other equipment. This paper investigates the various sources of error in the measurement of the guidance angle. An effort is made to separate the errors introduced by the measuring system from those generated by the antenna and transmitting equipment.

The equipment consists of a Signal Analyzer which computes the guidance angle using the log-video output of a standard MLS receiver, a guidance angle simulator, the antenna under test mounted on a controllable turntable, an optical encoder to measure the actual position of the turntable, and a receiving antenna mounted on a tower. These equipments are located in a test range which is optimized to reduce extraneous factors like multipath effects.

Errors in the measuring system include the accuracy of the location of the phase centre of the MLS antenna, nonlinearities and other errors in the receiver, calibration of the test range, and temperature effects. The MLS antenna contributes errors through the skewing of the beam, especially for large guidance angles, as well as errors introduced by the conical scan which results from digital beam steering. In addition, there may be errors introduced by failures in the phase shifters, and the correctness of the algorithm used for beam steering. The intent of the antenna test is to isolate and quantify these last errors, thus providing a method for judging the performance of MLS antenna and transmitting equipment.

INTRODUCTION

The Microwave Landing System (MLS) has been chosen to replace the Instrument Landing System (ILS) which has been in use at the world's major airports for more than 40 years. It is expected that, within ten years, MLS equipment will be operational at a majority of Canada's airports, as well as at airports in the U.S. and in other parts of the world. The safety, and efficiency, of airport operations will depend on the reliability and accuracy of the installed MLS equipment. The ability to maintain the required accuracy will in the first instance depend upon the avionics, especially MLS receivers, on board the aircraft. However, the ultimate consideration in determining the accuracy and reliability of MLS systems will be the MLS transmitting systems installed at the ends of the runways.

Prior to the widespread introduction of Microwave Landing Systems, it is necessary to develop methods and equipment that can be used to test, certify and commission MLS transmitting equipment. Furthermore, test equipment is needed for in-service testing of the MLS transmitters. This paper describes the development of a reliable and versatile test set that can be used for a first article test for any MLS

Azimuth (AZ), Elevation (EL) or Back Azimuth (BAZ) antennas.

In order to certify MLS transmitter equipment, a number of parameters need to be studied. These include radiated power levels, sidelobe levels, timing of various signpost signals, the timing of differential phase-shift key (DPSK) data, and the accuracy of the time-reference scanning beam of the MLS transmitting system. This paper deals primarily with the last item.

THE TEST RANGE AND EQUIPMENT

In order to evaluate the performance of MLS antennas and transmitting systems, a test range has been developed by Transport Canada at the David Florida Laboratory (DFL). The system configuration for field trials on the MLS Signal Analyzer is shown in Figure 1. The antenna to be tested is mounted on a turntable, fitted with an optical encoder. This is installed on the roof of DFL. Beside the turntable is a hut, which houses the SA4116 positioner, and the encoder interface. A tower is located about 400 m away, on which the MLS receiving antenna is installed. The configuration was carefully chosen to minimize multipath and reflection from adjacent structures to enable first article testing of MLS antennas under near-ideal conditions. At the foot of the tower is a hut, which houses the Signal Analyzer, the receiver and the SA2012A turntable controller. The SA2012A controller was linked to the SA4116 positioner through IEEE-488 extenders. The encoder interface in the roof hut communicated with the Signal Analyzer through a 2400 baud RS232 serial link.

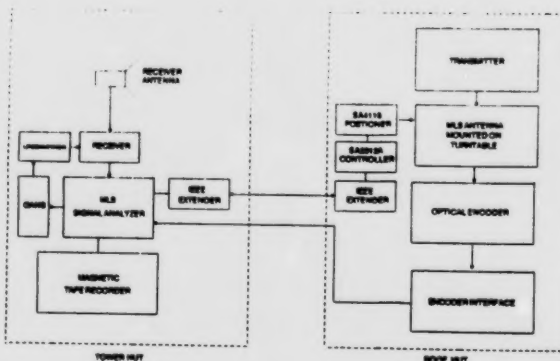


Figure 1 The Test Equipment at DFL.

The Signal Analyzer comprises a microprocessor, a 12-bit digitizer, a dynamic memory access (DMA) controller, and a guidance angle simulator (GANS). The digitizer converts the analog log-video from a commercial MLS receiver into 12-bit samples. These are input to the processor through the DMA controller and processed in software. The GANS can be used to generate Taylor-weighted antenna beams to simulate the real MLS signal, or to play back a previously recorded MLS log-video tape.

GUIDANCE ANGLE ACCURACY

One of the major functions of the Signal Analyzer is to determine the accuracy of guidance angle measurements. In order to do this, it is necessary to analyze all sources of error, and devise a means of determining their contributions to the total system error. The accuracy of guidance angle measurement achieved by the Signal Analyzer depends on a number of factors. These include:

- 1) The correctness of the algorithm used.

The algorithm used consists of finding the beam centres for the TO and FRO beams at levels with respect to the peak from -1dB to -10dB in 1 dB steps. The -3dB centres are then found by linear interpolation of the observed values. The time difference between the -3dB centres is proportional to the guidance angle. The angle derived from a single frame gives us a probable error of 2 microseconds (the resolution of the measurement). This corresponds to an angle of 0.04° . The use of 10 points to obtain the -3 dB beam-centre introduces an improvement of at least $\sqrt{10}$ in the guidance angle accuracy. Sixteen frames are used to obtain a single guidance angle, thus improving the accuracy by another factor of 4. Therefore the expected error in guidance angle measurement is $(0.04/\sqrt{10})/4 \approx 0.003^\circ$.

- 2) Non-linearities in the receiver.

Non-linearities in the receiver could affect the guidance angle measurements. The effects depend on the shapes of the TO and FRO beams, as well as on the nature of the non-linearities. The receiver non-linearity could be different for increasing signals and decreasing signals, causing a change in the shape of the beams. If we first consider small guidance angles near boresight, we can assume that the TO and FRO beam shapes are identical, and that the beams are not skewed. If this is the case, non-linearities in the receiver will distort both the TO and FRO beams identically. Hence, the -3dB centres would be separated by the same time, as if there were no distortion, and no error will be introduced in the angle measurement. However, the midpoint of the scan could be shifted, causing an error in the asymmetry.

For the case of large guidance angles, far from boresight, the TO and FRO beams could be skewed. This means that, if the non-linearity of the receiver shifts the -3dB point of the TO beam centre by a certain amount in time, the -3dB point of the FRO beam centre would shift in the opposite direction, as shown in Figure 2 below. This shift in the -3dB beam centres causes an error in the angle as seen by the Signal Analyzer.

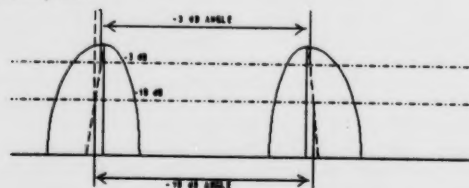


Figure 2 Effect of Non-linearities in the Receiver

Let us represent the error by $\delta\theta$. This error is proportional to the skew of the beam. The skew or tilt of the beam is measured by the slope of the beam-centres expressed in degree/dB. The error $\delta\theta$ is given by

$$\delta\theta = \delta p \cdot m$$

where δp is the error in estimating the -3dB beam-centre, and m is the slope of the beam-centre line in degree/dB. It is easy to see that δp is determined by the non-linearity in the receiver, particularly if the signal saturates the receiver. It is therefore extremely important to ensure that saturation does not occur. The tilt, m , is expected to vary with guidance angle, increasing as the angle moves away from boresight. It will contribute most to the angle error at large guidance angles. A fairly pessimistic estimate of m is one-half the slope of the beam at the -3dB point which is of the order of 0.15 degree/dB. Thus, m can be estimated to be 0.08 degree/dB. A very pessimistic value for δp is 1%, due to a 1% error in estimating the peak of the beam. At the -3dB point, this error is estimated to be 0.03 dB. Using these, we get

$$\delta\theta = (0.03 \cdot 0.08) \approx 0.0024^\circ$$

The above error is introduced by non-linearities in the receiver, acting on asymmetric beams.

- 3) Random or asynchronous noise introduced by the receiver.

It is possible that the receiver can, through components such as switching power supplies, superimpose an asynchronous noise on top of the log-video. CAL measured this noise, and found that there is a component of this noise at a frequency of 25 kHz. However, the magnitude of this noise component was extremely small, amounting to only -60 dB.

The effects of random noise can be estimated by performing statistical analysis of a number of frames digitized from the same input signal. This can be done using the GANS as the signal source. A tape was generated for two angles, for each of the four functions (AZ, BAZ, HAZ and EL). Sixteen frames were recorded on tape for each of the functions. The GANS was then used to play back one of these frames repeatedly. The digitized data was then analyzed for a total of 5 scans (or 80 frames). The root mean square deviation of each frame with respect to the first is a measure of the noise in the system. In addition, the mean angle, and the standard deviation of the measured angle were also computed. The mean is the best estimate of the angle from the available data. This would reveal any offset or bias (systematic error) in the measured value. The probable error was estimated assuming that the random errors were normally distributed. The probable error (PE) is defined as that value for which there is a 50% chance that the measurement will be between the values $X - PE$ and $X + PE$, where X is the expected value. For a normal error distribution, the probable error is given by 0.674σ . Thus, this analysis gives us a measure of both systematic and random errors. It was found that the measured angle was close to that obtained from the statistical analysis. The agreement was always better than 0.005 degrees.

- 4) Signal level distortions introduced by non-linearities in the A/D converter.

Non-linearities in the A/D converter will not significantly affect the shape of the log-video signal.

- 5) The accuracy and stability of the 2 microsecond clock used by the digitizer, and quantization errors introduced by the conversion process.

Assuming a beam slope of 0.15 degree/dB, an error of 0.001° in the guidance angle is generated by a jitter of 50 nanoseconds in the clock. Thus, stability of the A/D clock is

not a significant source of error, since stable crystals are readily available.

6) Range Calibration Errors.

The conical scan correction is sensitive to the elevation angle. Therefore, it is essential to know the geometry of the range accurately. The phase centre offset corrections, in particular, are sensitive to even very small errors in the distances measured. The allowed error in range measurement is given by

$$\frac{\Delta R}{R} = \frac{\tan \theta}{\tan \theta'} - 1$$

In the above, R is the horizontal range from the receiver to the phase centre of the antenna, and ΔR is the error in the measurement of R. Thus, an error of 0.005° in the guidance angle corresponds to an error of just 7 cm on a 360 m range. In the range tests, the phase centre of the antenna was assumed to coincide with its geometric centre. An equation similar to the one above applies to the left-right displacement of the antenna phase centre. Another range parameter that needs to be measured accurately is the height of the receiver antenna above the phase centre of the antenna under test. The elevation angle, related to this parameter, affects the measured guidance angle due to the conical scan pattern generated by MLS antennas. The conical scan error is given by

$$\sin \theta' = \frac{\sin \theta}{\cos \phi} \quad (1)$$

where θ' is the true guidance angle; θ is that measured; and ϕ is the elevation angle. The equation above is valid at large distances from the MLS antenna, where the phase centre offsets are negligible. The errors introduced by the conical scan are shown in Figure 3, for various elevation angles. An error of 10 cm in the height introduces an angle error of 0.005° at a guidance angle of 60°, if the elevation angle is 10°. At an elevation of 20°, the allowed error in the height is only 5.6 cm. Thus it is clear that accurate range measurements are necessary to measure guidance angle accuracy.

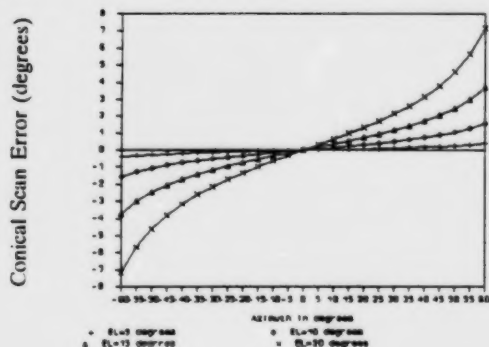


Figure 3 Guidance Angle Error due to Conical Scan

(7) Beam Skewing at Large Angles

However, there is another source of error that has so far not been considered. This arises from the fact that the guidance angle has been arbitrarily defined by ICAO to be measured at the -3dB level. The level at which it is defined is inconsequential, provided the beams are not skewed. If, however, they are, as is the case far from boresight, the errors could be large. The measurements at DFL yielded a worst-case tilt of 0.01 degree/dB. This corresponds to a guidance angle error of 0.05°, which is 10 times the targetted accuracy. The error discussed here is quite different from that described in section (2) above, where the error was due to non-linearities in the receiver. The beam skew exhibited by a commercial MLS AZ antenna tested at DFL was analyzed using transport Canada's VAX computer. The difference between the angles computed at -1dB and -10dB was found to be as large as 0.032° at a guidance angle of 50°, and -0.022° at -50°. The shapes of the TO and FRO beams, clearly showing the beam skew, are illustrated in Figure 4 below. It was found that the error in the angle could be written as

$$\delta \theta = (t(\text{TO}) - t(\text{FRO})) \cdot v \cdot \text{SF} / 2 \quad (2)$$

where SF is a skew factor estimated at 2.5, $t(\text{TO})$ and $t(\text{FRO})$ are the beam centres as computed by the signal analyzer (as defined by ICAO), and v is 0.02 degrees/microsecond.

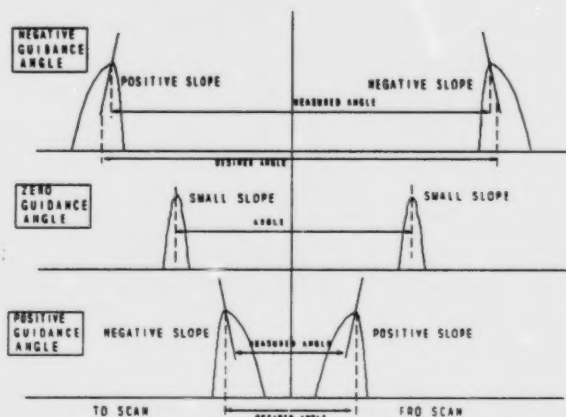


Figure 4 Observed Beam Skew

The data collected at the David Florida Laboratory (DFL) have been analyzed to study the effects of the tilt in the TO and FRO beams. The scan angles have been computed using the beam centres at the -1 dB through -10 dB levels. These were calculated for five azimuth angles from -50 to +50 degrees, applying only the conical scan correction, but excluding the phase centre offset corrections. It was observed that the value of the computed guidance angle changed with the level, consistent with the observance of the beam tilt. Further, the variation in the angle with level increased with the angle from boresight. The uncertainty in the angle, or spread in the error, was about 0.05 degrees at -50, 0.04 at -30; 0.02 at 0; 0.04 at 30; and 0.09 at +50 degrees. The results for the commercial MLS antenna, where the curves show erratic jumps, indicate that the approach of performing a linear regression may not be optimal. The erratic behaviour seems to suggest that the DFL range may be subject to some multipath or other effects. Apart from these peculiarities, larger errors are evident. For instance, at -50° azimuth, a correction of -1.2° had to be applied to bring the angles closer

to the nominal value. Figure 5 shows the results obtained for a guidance angle near -50° , one curve for the first frame, and another for the average of 16 frames.

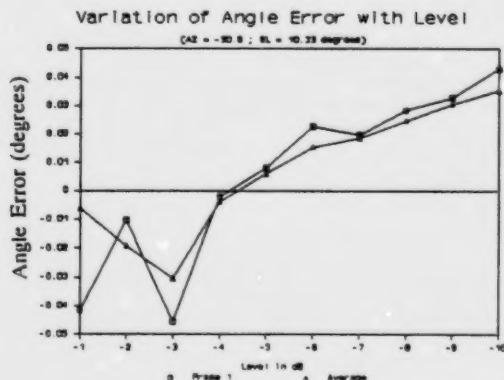


Figure 5 Angle Error due to Beam Skew, AZ = -50°

There are two effects involved in these results. The first is the variation in angle from level to level. This is caused solely by the tilt in the TO and FRO beams. The second effect is the offset. An offset error can result from a number of causes. Firstly, the corrections for the conical scan and the phase centre offset may not have been implemented correctly. This possibility is rather remote, since both Transport Canada and CAL have studied the algorithm and its implementation very carefully. However, even though the algorithms may be correct, the input parameters may not have been. It is clear from Figure 3, which shows the effect of conical scan alone, that the error is sensitive to the elevation angle. Any error in this could cause large errors in the computed angle. The same is true of the phase centre offset correction. Thirdly, the offset error could be due to errors in the optical encoder. Although this is a very accurate instrument, any slop in the mounting can cause large errors in the angle measured. The results of the DFL tests show that, once all sources of offset errors are removed, an uncertainty of about 0.1 degree still exists. This is the error due to the tilt of the TO and FRO beams. Further, this error could vary from antenna to antenna, depending on the algorithm used for beam steering. The results indicate that method used for computing the guidance angle needs fine tuning. The possible new approaches include using an average of 16 frames, extrapolating the beam centres to 0 dB rather than 3 dB, or using a formula such as equation (2) above.

In summary, the single largest contribution to the guidance angle inaccuracy arises from the conical scan. This can be corrected in operational receivers using the formula given in equation (1), using the measured azimuth and elevation angles. It is particularly important to do this for curved approaches. The residual error is then of the order of 0.1 degree, which can be removed for calibrated antennas. As the aircraft approaches the runway, the phase centre offset corrections become more important. These errors can be compensated in range measurements, as was done for the DFL tests, but cannot be corrected in operational systems.

CONCLUSIONS

A reliable and versatile test set, comprising an MLS signal analyzer and a guidance angle simulator, has been described in this paper. The equipment is an essential element in testing and certifying MLS installations. It will be used for first-article testing of MLS azimuth and elevation antennas. It will also be utilized to certify and commission MLS transmitting equipment. The analyzer has the following features:

- has a targetted accuracy of $\pm 0.005^\circ$ which is an order of magnitude better than commercially available receivers,
- produces reliable and repeatable results,
- permits Transport Canada to investigate the accuracy of guidance angle beams output by MLS transmitters, and
- provides a means for recording the actual log-video of the transmitted signal.

The accuracy to which various range measurements need to be made has been investigated. The effect of beam skewing on the guidance angle exhibited an error of as much as 0.1° at a guidance angle of 50° , for an azimuth antenna tested at DFL. The error in the guidance angle measurement due to the conical scan was also investigated, and found to be as large as 2° at a guidance angle of 50° at an elevation of 15° . The error due to the phase centre offset will be the subject of future study.

Accurate guidance angle measurements can be made at small angles using the signal analyzer described in this paper. At larger angles, sources of error were identified that would not have been found without the use of the signal analyzer and the guidance angle simulator. Accurate guidance angle measurements at large angles requires the minimization of these errors by range calibration, and modification to the guidance angle measurement algorithm.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

Image-Supported Navigation for Testing Instrument Landing Systems

U. Brokof and F. Knabe

German Aerospace Research Establishment (DLR)
Institute for Flight Guidance

P.O. Box 32 67

D-3300 Braunschweig

Federal Republic of Germany

Abstract

The requirements of the International Civil Aviation Organization (ICAO) as well as economic and safety considerations put restrictions on testing of Instrument Landing Systems (ILS) at airports with a high traffic load. The Cooperative Flight Measurement System Group (GFMS) develops under these aspects with the support of DLR such a new testing system. It is based on on-board sensors only as inertial navigation system (INS), DME receivers and a video camera. High precision of flight path reference results is achieved by integration of INS data and video updates in a Kalman-filter with backward filtering. Thus calibration of category III ILS will be possible. As video updates are generated by automatic pattern recognition of runway thresholds, only one successful approach is necessary in principle for testing an ILS. The update technique and backward filtering enable measuring of non-runway-centred glide paths, as they are flown with future Microwave Landing Systems (MLS). Concepts, simulations and first results are discussed.

1. Introduction

Regular testing of radio navigation aids on their precision is an important contribution to safety in air traffic. International Civil Aviation Organization (ICAO) fixes the requirements on flight and ground inspection. BFS as the German Civil Aviation Authority manages civil radio navigation aids. GFMS (Cooperative Flight Measurement Group), an organizational unit of BFS, supervises civil and military installations with reference to ICAO requirements.

The GFMS testing equipment for ILS presently consists of an aircraft and a mobile Infrared Flight Inspection System (IFIS) as direction indicator on ground [4]. In spite of its high precision the system has still disadvantages resulting from the measuring principle of IFIS:

- Expense of ground staff and equipment,
- limited angular view (direction to be tested $\pm 1^\circ$),
- strong dependence on weather conditions,
- separate approaches for localizer and glide path.

These disadvantages are eliminated by application of an INS, which continuously produces position data in geographical coordinates. Directions to be calibrated such as localizer azimuth are derived from the INS position data and the coordinates of transmitter and receiver. Moreover computation of geometric distances to arbitrary points is possible. The drift-rate of INS position components is compensated by updates from other sensors. Multiple DME measurements are one way to update an INS. The accuracy is in the range of 30m - 200m (1%), dependent on available DME ground stations [1].

Better accuracy of updates can be achieved by using a video sensor. This sensor and a connected image processing system recognize outstanding landmarks with high accuracy. At airports threshold markings on runways are very simple but characteristic patterns which can be detected in real time by image processing today. As these patterns are the same on all airports

this kind of autonomous update is correspondingly applicable. In addition to that the sensor works nearly weather independent at the low altitudes of the flight tests.

Therefore GFMS decided to build up an ILS testing system with the following configuration: Basic navigation is done by an INS. Multiple DME improves en-route navigation and additional video updates yield high accuracy during approaches. As video updates are only point- but not direction-dependent like IFIS is, approaches with different headings are possible. This feature is especially important in regard to future testing of Microwave Landing Systems (MLS).

The Institute for Flight Guidance of the German Aerospace Research Establishment (DLR) has solved a similar task in building up the Avionics Flight Evaluation System (AFES). It was designed for generating highly precise flight trajectories as a base for calibrating new navigation techniques and systems. Therefore DLR received the commission to design the GFMS system, to develop data processing algorithms and to consult GFMS during tests until successful operation.

2. Requirements and basic ideas

The first part of this chapter summarizes the main ICAO requirements on ILS calibration and the corresponding economic aspects. The second part briefly presents AFES of DLR and describes common and different aspects of the GFMS system and AFES.

2.1 Requirements on accuracy and economy of testing ILS

Requirements of ICAO on testing of radio navigation aids are fixed in [4]. Allowable tolerances of localizer and glide path signals are given in the respective categories. A distance of 4000m between localizer antenna and runway threshold with a glide path of 3° is the basis of a typical example. For category III a tolerance in localizer course alignment of $\pm 0.01^\circ$ and a tolerance in glide path alignment of $\pm 0.025^\circ$ are derived. These values are adequate to an accuracy of the reference trajectory of 0.70m in lateral and 0.13m in vertical directions at runway threshold.

Economic aspects demand for minimizing of time and resources:

- Minimizing number of approaches,
- reduction of ground staff or equipment as far as possible,
- availability of calibration data shortly after test flight.

2.2 The Avionics Flight Evaluation System (AFES)

AFES was built up by DLR Institute for Flight Guidance in order to generate very accurate reference trajectories for testing of navigation systems [1, 6]. Main components are an INS as basic navigation unit, a DME-receiver, a microwave tracking radar and a lasertracker. AFES can be run in three configurations

dependent on accuracy demands. The accuracies achieved (1σ) are the following (distance 10 km, DME means multiple DME):

Configuration	Direction	Distance
INS/DME	0.15°	30m
INS/DME/Radar	0.01°	2m
INS/DME/Radar/Laser	0.005°	0.3m

Figure 1 gives an overview of AFES components and results.

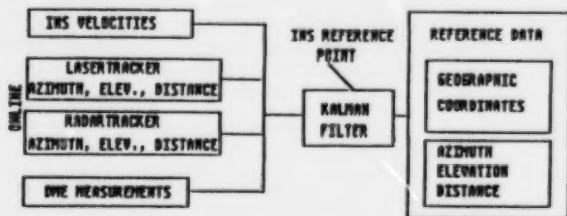


Figure 1 AFES measurement data processing and output

The accuracy which can be achieved by configuration 3 of AFES is in the same range as the accuracy required for an ILS testing system. A major difference is the update rate. Lasertracker measurements are used nearly continuously while the video update is only available at the end of an ILS test-flight. Therefore the corresponding algorithms have to be modified in order to achieve similar results.

3. Concept in detail

The GFMS concept of an ILS testing system is derived from AFES. Multiple-DME measurements serve as basic updates of the INS on en-route flights. INS and DME data are fitted together in an online Kalman-filter. Video updates are taken in order to improve reference data. As video updates with runway thresholds only occur at the end of an approach, offline backward-filtering is effected to extend the accuracy of the video updates widely back into the approach trajectory. Figure 2 gives an overview of GFMS system concept on measurement data processing, where IFIS data are integrated to allow a sliding change from the old to the new system.

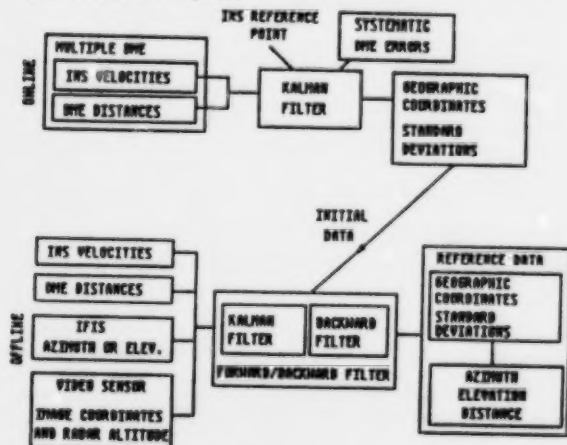


Figure 2 GFMS system concept of measurement data processing

3.1 Filter algorithms

The Kalman-filter algorithm is separated into a prediction part and an update part. The update part will only be activated if measurement data are available.

Raw errors in the measurement data are detected by an identification process. Within the filtering algorithm a simple method, which reduces computation time considerably is used in order to avoid zero multiplications in connection with sparse matrices (transition matrix) [2].

All measurement data processed in the past influence the accuracy of state variables and therefore of navigation. Measurements, which will occur in the future, do not have any influence on current navigation results. In offline data evaluation one can make use of the past and the future in the flight path computation at a certain instant of time. This is the basic idea of the RAUCH-TUNG-STRIEBEL smoothing algorithms [5]. It requires data storage along the respective part of the trajectory.

3.2 INS error model and observation equations of DME and IFIS

The state vector consists of 13 components, 7 of them relate to INS error model and 6 to the DME error model:

- $\Delta\lambda, \Delta\phi$ position error components in geographical longitude and latitude,
- $\Delta\lambda, \Delta v_y$ errors of INS velocity components in east and north,
- α, β, ψ alignment errors of INS,
- $\Delta\rho_{DMEi}$ DME bias of $i = 1$ to 5 different DME transmitters,
- $\Delta\rho_{RECV}$ DME bias of on-board receiver.

Five DME-stations are used in a certain period of flight time. They are interrogated sequentially in intervals of two seconds. A certain DME station is replaced by a new DME station if this comes in closer range. Observation equation for DME measurement is

$$\begin{aligned} \Delta z &= \rho_{DMEi} - \rho_{NAV} \\ &= \left(\frac{\partial \rho}{\partial \lambda} \right)_{NAV} \cdot \Delta\lambda + \left(\frac{\partial \rho}{\partial \phi} \right)_{NAV} \cdot \Delta\phi + \Delta\rho_{DMEi} + \Delta\rho_{RECV} + r_{DMEi} \end{aligned}$$

Index NAV means values computed at current and not updated navigation position. r_{DMEi} is the high-frequency part of DME error. Its variance σ_{DME}^2 is estimated to $(100m)^2$.

For the update of the reference flight path with IFIS measurement data the following observation equation is used.

$$\Delta z = \rho_{IFIS} - \rho_{NAV} = \left(\frac{\partial \rho}{\partial \lambda} \right)_{NAV} \cdot \Delta\lambda + \left(\frac{\partial \rho}{\partial \phi} \right)_{NAV} \cdot \Delta\phi + r_{IFIS}$$

In this equation r_{IFIS} represents the random error of IFIS measurements. Its variance σ_{IFIS}^2 is assumed to be $(0.02^\circ)^2$.

Differential quotients in equation (1) and (2) are derived in detail in [2, 6].

Exact computation of Δz is decisive for precise results of filtering.

3.3 Observation equation of video sensor measurements

Original video sensor measurements are the pixel coordinates x_{PIX} and y_{PIX} of the image of intersection M between runway center line and the backward edge of threshold pattern (Figure 3). The image processing system uses the high contrast and the significant pattern of runway threshold for recognizing point M. In flight direction the end of the stripe pattern is detected. Lateral detection of point M is enabled because the center gap between the stripes is nearly twice as wide as the other gaps.

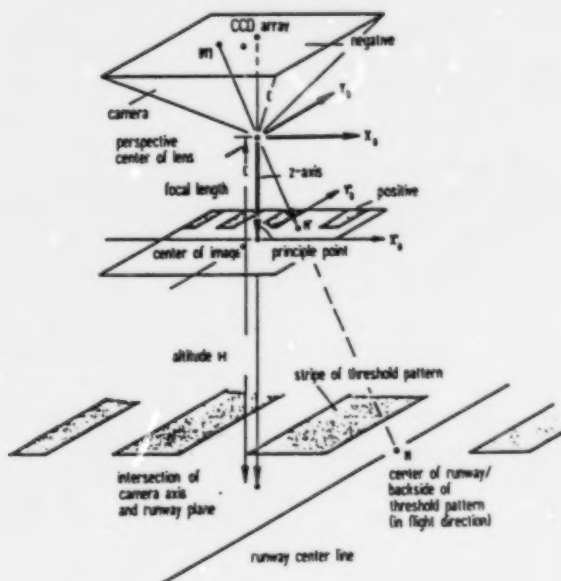


Figure 3 Geometric configuration of video sensor measurement

The original measurements have to be transformed into a coordinate system in which they can be compared with measurement data of INS in a linear relation.

First step is the scaling of x_{PIX} and y_{PIX} . GFMS uses an AQUA-TV video sensor with FUJINON lens. Its aperture ratio is 1:1.6, focal length is 5.5 mm and aperture angle $50^\circ \times 40^\circ$. Thus a typical flight altitude of 15m leads to a covered ground area of $14m \times 11m$. In comparison to this the dimensions of runway threshold pattern at airport Braunschweig are $30m \times 30m$. The scaling factors s_x and s_y are derived from the number of used pixels (here 604×287) and length and width of CCD array (here $6.0mm \times 4.5mm$). Results are the image coordinates x_B and y_B of M':

$$x_B = s_x x_{PIX} = \frac{6.0mm}{604} x_{PIX}$$

$$y_B = s_y y_{PIX} = \frac{4.5mm}{287} y_{PIX}$$

Definition of image coordinate system B (Figure 3) is the following: x-axis and y-axis are aligned to the direction of rows and columns of the CCD array, z-axis is directed from the principle point of the array as coordinate origin to the exit pupil of lens. With coordinates x_0, y_0 of principle point and focal length c this leads to

$$\begin{aligned} x_B &= x_0 \\ y_B &= -(y_0 - y_0) \\ c \end{aligned}$$

The next step is a transformation into a strapped down coordinates system F. Three rotation angles α, β and κ with a particular definition [2,3] are used:

$$\underline{X}_F = \underline{R}_K(\alpha, \beta, \kappa) \underline{X}_B \quad \underline{R}_1 \dots \text{Rotation matrix}$$

The orientation of the strapped down system F against the geodetic system G with its axes x pointing north, y pointing east and z pointing down is achieved by rotations with roll, pitch and heading angles (θ, ϕ and ψ)

$$\underline{X}_G = \underline{R}_F(\theta, \phi, \psi) \underline{X}_F$$

The coordinate system used for the comparison between predicted position and video measurement is a right handed terrestrial system T with origin in M. Direction of z-axis is zenith, y-axis is north-oriented.

Factor m_M is introduced for scaling:

$$\underline{X}_T = -m_M \cdot \underline{R}_T \cdot \underline{X}_G = -m_M \cdot \underline{R}_T \cdot \underline{R}_F \cdot \underline{R}_K \cdot \underline{X}_B$$

$$\underline{R} = -\underline{R}_T \cdot \underline{R}_F \cdot \underline{R}_K = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

$$\begin{aligned} \underline{X}_T &= \underline{X}_T = m_M \cdot \begin{pmatrix} r_{11} \Delta x - r_{12} \Delta y + r_{13} c \\ r_{21} \Delta x - r_{22} \Delta y + r_{23} c \\ r_{31} \Delta x - r_{32} \Delta y + r_{33} c \end{pmatrix} \\ \underline{Y}_T &= \underline{Y}_T = m_M \cdot \begin{pmatrix} r_{11} \Delta x - r_{12} \Delta y + r_{13} c \\ r_{21} \Delta x - r_{22} \Delta y + r_{23} c \\ r_{31} \Delta x - r_{32} \Delta y + r_{33} c \end{pmatrix} \\ \underline{Z}_T &= \underline{Z}_T = m_M \cdot \begin{pmatrix} r_{11} \Delta x - r_{12} \Delta y + r_{13} c \\ r_{21} \Delta x - r_{22} \Delta y + r_{23} c \\ r_{31} \Delta x - r_{32} \Delta y + r_{33} c \end{pmatrix} \end{aligned}$$

with $\Delta x = x_B - x_0$ and $\Delta y = y_B - y_0$.

Z_T is the altitude H of the perspective center above runway and can be derived from the radar altitude H_R . Thus the scaling factor m_M is

$$m_M = \frac{H}{r_{31} \Delta x - r_{32} \Delta y + r_{33} c}$$

The relative positions of the radar altimeter, the video sensor and the INS on-board the aircraft are considered [3].

As Kalman-filtering with UD-factorization is used [2, 6], only non-correlated measurement components may be processed successively. Thus observation equations have to be transformed into coordinate system with its axes in longitudinal and lateral runway directions. This leads to the final observation equations

$$\Delta z_{lat} = (R_E \cos \phi \cos \psi_0) \cdot \Delta \lambda + (-R_N \sin \psi_0) \cdot \Delta \phi + r_{lat}$$

$$\Delta z_{lon} = (R_E \cos \phi \sin \psi_0) \cdot \Delta \lambda + (R_N \cos \psi_0) \cdot \Delta \phi + r_{lon}$$

where R_E and R_N are radii of earth curvature in east and north and ψ_0 is runway direction. These equations can be used for a video update as well as for a manual update by the operator who observes the threshold overflight.

The baro-inertial flight altitude is corrected by using temperature measurements during the approach and a radar altimeter measurement at runway threshold.

4. Simulation, implementation and first results

The offline navigation program has been tested by simulations with position and ILS data of Nuremberg Airport. Different configurations have been tested for an approach of 50 km distance, 5000 ft altitude and 70 m/s velocity. INS, DME, IFIS and video errors were simulated. The results of simulation are plotted in a stereographic projection so that deviations between reference position and extended runway center line are shown directly (Figure 4 and 5).

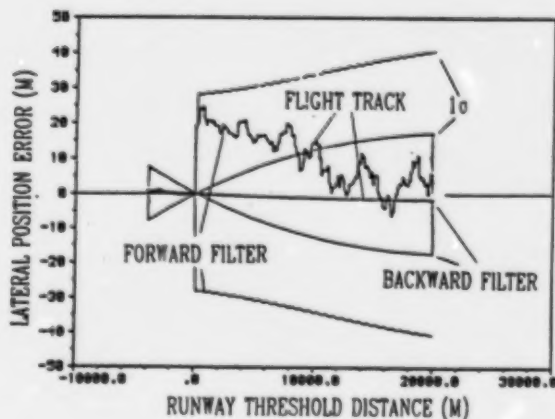


Figure 4 Simulated approach with multiple DME and one video update

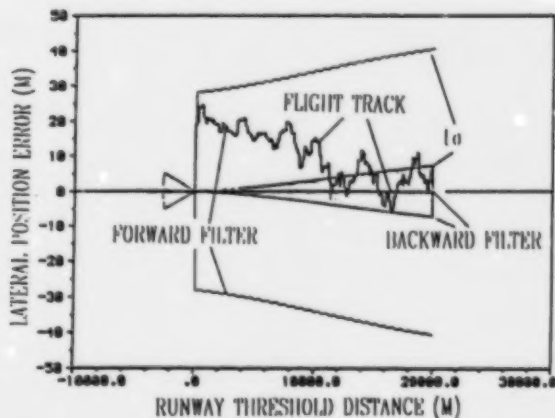


Figure 5 Simulated approach with multiple DME and two video updates

The testing of ICAO category III ILS requires two video updates.

After the simulation the software written in FORTRAN was implemented into the on-board-computer (with 68020 processor) by a software company.

First test flights, carried out in June 1989 at Hanover Airport verified the proper running of the complete system. The first flight results also show that the achieved accuracy seems to be sufficient for testing ILS category III installations.

5. Conclusions and outlook

The new ILS testing system of GFMS uses video updating of an INS to achieve calibration accuracies required by ICAO. A video sensor automatically recognizes characteristic stripe patterns of runway thresholds. This technique enables the system to work completely without ground staff or special installations.

As video updates occur only at the end of an approach, backward-filtering is used for improving the accuracy during the approach. Computation of calibration data is accomplished in-flight a few minutes after the approach. After that the system is ready for further test flights.

Simulations have shown that the required accuracy for category III calibration flights can be achieved with two video updates at both ends of the runway. First test flights at Hanover Airport confirm this.

As three-dimensional position data are computed continuously, approaches at different approach angles or curved approaches are possible. Thus the system is also ready for future MLS testing.

The comparison of the GFMS ILS testing system with AFES reference system, will be done later in order to prove the accuracies achieved in the simulation and the first flight test.

References

- [1] K.H. Hurraß and P. Schulz, Das Avionik-Flugerprobungssystem (AFES). DFVLR-Mitteilung 85-01, Köln 1985
- [2] U. Brokof, Das Navigationsprogramm für die Referenzdatenerzeugung der Gemeinsamen Flugvermessungsstelle (Offline-Version). IB 112-89/11 DLR, Institut für Flugführung, Braunschweig 1989
- [3] U. Brokof and H. Schwind, "Referenzbahnbestimmung für die Überprüfung von Flugsicherungsanlagen durch Bildstützung mit Hilfe einer CCD-Kamera" in Bildsignalverarbeitung für Aufgaben der Flugführung. DFVLR-Mitteilung 88-32, Köln 1988
- [4] Manual on Testing of Radio Navigation Aids- Volume II: ILS (Instrument Landing System). International Civil Aviation Organization, DOC 8071, Third Edition, 1972
- [5] A. Gelb, Applied Optimal Estimation. The M.I.T. Press, London 1974
- [6] U. Brokof and H. Haensel, Das Referenzbahnsystem des Avionik-Flugerprobungssystems (AFES). IB 112-86/11, DFVLR, Braunschweig 1986

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

COMPREHENSIVE MLS TESTING FOR CATEGORY I, II & III OPERATION

Alex E. Smith, MSI Services, Inc.
600 Maryland Ave, S.W., Suite 440, Washington D.C. 20024, U.S.A.

Abstract

This paper reviews the test requirements necessary to ensure that an installed Microwave Landing System (MLS) will meet all of the exacting requirements to provide for category I, II, and III operations. General requirements for MLS testing are presented, and testing is considered throughout the entire life cycle of the equipment, from design to service. In evaluating the requirements of a comprehensive MLS test program, the top-down processes of specifying requirements and designing to requirements are applied to system testing. Resulting from this, an overview is given of a master test plan for MLS which spans the entire system life cycle. The paper is written primarily from the perspective of the aviation agency. Guidance is given for the agency to design a master test plan which is rigorous but also practical for implementation by the MLS contractor.

Introduction

The MLS comprises an Azimuth equipment, Elevation equipment, and a Distance Measuring Equipment (DME). MLS is required to provide high integrity precision guidance throughout a large volume of space. As a result the testing requirements for MLS are more stringent than those of any other radio navigation aid. This paper reviews the test requirements necessary to ensure that an installed MLS has met, and will continue to meet throughout its operational life, all of the exacting requirements to provide for category I, II, and III operations. The paper is written from the perspective of the aviation agency, as ultimately the agency is responsible for the operation or approval of the installed system. This applies worldwide, and in this paper testing requirements are evaluated from a European and a North American perspective. A master test plan which will allow the agency to perform comprehensive MLS testing to satisfy technical and operational requirements is discussed. Particular attention is given to the use and suitability of existing standards and documentation which are available to agencies and contractors to design and structure the master test plan.

MSI Services Inc., is a multi-disciplinary engineering and management services firm, and since its formation in 1976 has provided support to Government agencies in major systems acquisitions programs, and in this role has worked closely with agencies and contractors. It is with the benefit of this experience that a comprehensive MLS test program is discussed which is rigorous for the needs of the agency, and practical for implementation by the contractor.

The Requirements For MLS Testing

The objective of a comprehensive test program is to ensure that a system meets the technical requirements to serve its operational purpose when installed. As is true for any system, the requirements for aviation navigation systems begin as operational requirements which are based on a fundamental operational need. The operational requirements are translated into technical requirements which in turn are developed into system requirements. Figure 1 shows the relationship between the operational need and the system requirements. This relationship can be used as a framework to plan for comprehensive MLS testing. By planning from the

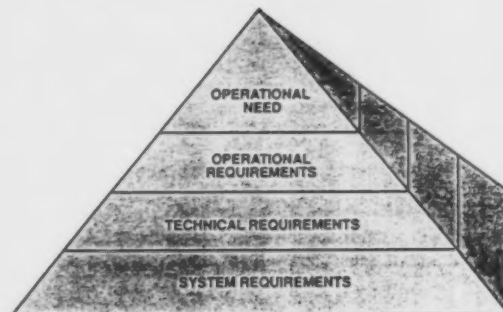


Figure 1. Hierarchy of Requirements

top-down, testing can be performed from the bottom-up to ensure the system meets all of the requirements.

Hierarchy of MLS Requirements

The Operational Need: The operational need for MLS arose as a result of the limitations of Instrument Landing Systems (ILS). It was recognized that ILS has a number of basic limitations: it is site sensitive and can be expensive to install; only 40 channels are available and frequency congestion is a serious problem; and it lacks the flexibility needed for future aircraft operations. It was these limitations that stimulated the search for a better precision landing system. In the late 1960's over 50 different systems, most of them using microwave technology, were proposed to overcome the limitations of ILS. To preclude the prospect of having many different types of approach systems in use in different countries the Radio Technical Commission for Aeronautics (RTCA) formed a special committee to recommend a replacement that could be endorsed by ICAO.

The Operational and Technical Requirements: Over the past 10 years ICAO panels have established the operational and technical requirements and criteria related to the introduction of MLS. The operational requirements for MLS are published in an ICAO report¹ and the technical requirements, or standards, are published in ICAO Annex 10 to the International Standards². It is from these technical requirements that aviation agencies derive system requirements.

System Requirements For MLS: System requirements for MLS are contained in agency procurement specifications. An agency procurement specification will usually take the form of a system/segment specification³ which will incorporate all of the relevant system performance requirements from the ICAO technical standards. This specification is the lowest level in the hierarchy of requirements for system performance. The system specification will define agency specific requirements such as: interface and communication requirements, remote maintenance and monitoring requirements, automatic fault isolation and detection, design and construction standards, and human engineering requirements.

Writing Testable Specification Requirements

As testing is performed from the bottom-up in the requirements hierarchy, system requirements are

evaluated here for testability. System requirements may be categorized as qualitative or quantitative, and most system specifications usually contain both. It is difficult for the agency to verify that the contractor has met all of the specified qualitative requirements as the agency and contractor may (will!) have different perceptions of these requirements. It could be argued that as qualitative requirements are difficult to verify, their incorporation into agency specifications is of limited use. A qualitative requirement is not one that is untestable, but one for which the test standard may be undefined. However, current practices and standards do allow the use of qualitative requirements⁴. It is good practice in specifying requirements to define qualitative requirements only when it is not feasible or practical to define quantitative requirements. Typically for MLS quantitative requirements include integrity, continuity of service, reliability, and maintainability; and qualitative requirements include frangibility, modularity, maintenance accessibility, and software integrity.

Test By Demonstration or Analysis Both qualitative and quantitative requirements can be tested by demonstration or by the use of analysis techniques. Demonstration is where a characteristic is capable of being measured. Analysis or prediction techniques are mathematical or statistical estimates of characteristics or attributes and are used where demonstration is not practical in terms of cost and time. Analysis techniques are used prevalently during the early part of the system design life cycle to verify the fundamental system design prior to starting the equipment prototype build. This should include a part/module and functional (including hardware and software functions) Failure Modes, Effects and Criticality Analysis (FMECA) and analyses of: reliability such as Mean Time Between Failures/Outages (MTBF, MTBO), integrity, environmental performance, and frangibility. Demonstration testing will be used more during the development stages of the system as the equipment build progresses.

The extent of the system testing will vary with the complexity and functionality of the system. A category III MLS system will require more testing than a category I or II system as it has additional functions associated with reliability and integrity, such as duality and end to end integrity checking. For functions which are common to category I, II, and III systems, there is no reason for the extent of testing to differ. This is because analysis techniques and demonstrations designed to satisfy category III systems (for agency procurements) will also meet category I and II requirements. Additionally, if any system category upgrade capabilities are envisaged, they should be part of the initial design and analysis process.

Testing Throughout The Design Life Cycle

The system design life cycle takes the system from initial design, through development and prototype stages to a state of readiness for production. For safety critical systems such as MLS, the agency requires more than the delivery of an end product. In order to verify and validate that the system will meet all of the exacting system requirements, particularly for category III MLS, the agency must have involvement in the development process. Bottom-up testing must be performed at different stages in the system design life cycle. The master test plan should determine and control the role of testing in the design and development process. This role should not be left to the contractor to define. The following discusses the

use of existing standards and documentation which are available for agencies to incorporate design life cycle testing in the master test plan.

In the U.S., system development and testing of Government major systems acquisitions is controlled by the rigid enforcement of U.S. standards. For MLS the standards are a combination of military and FAA documents. Combining the military standard for system technical reviews and audits⁵ and the FAA standard for electronic equipment quality assurance⁶, as shown in Table 1, illustrates the role played by testing in the design process. Table 1 represents a summary of formal agency review requirements and is simplified in this paper to illustrate the role of testing. The use of U.S. military standards is widespread in aviation and it is common practice throughout North America and Europe to use some U.S. standards in equipment specifications. Similar to the FAA, many agencies publish their policies in standards for electronic equipment quality assurance.

Table 1 The Role of Testing in the Design Process

<u>Contractor Design Reviews and Tests</u>	<u>Agency Test Requirements</u>
•System Requirements Reviews	
•Software Specification Review	
•Preliminary Design Review	•Hardware and Software Configuration Item Test Plans Available
	•Quantitative Analyses of Reliability, Maintainability, Integrity, FMECA.
•Critical Design Review	
•Test Readiness Review	
	•Hardware and Software Configuration Item Functional and Physical Tests
•Contractor Preliminary Tests	•System Prototype Performance Tests
•Design Qualification Tests and Reviews	•Full System Design Tests
	•System Performance Testing Over Service Condition Range
	•Reliability Demonstration
•Production Readiness Review	

Each of the reviews is performed at a different stage in the equipment design life cycle. The critical design review is conducted for each configuration item (board, unit, and module designs) when the detailed design is essentially complete. As hardware and software are not required to exist at the critical design review stage the only testing that can

be performed prior to this review is testing by analysis. Following the critical design review, individual configuration items should be tested for functionality and the product baselines should be established. These items are tested for compliance with contractor prime item specifications which are developed from the agency procurement specifications, for configuration items below system level.

The purpose of the contractor preliminary tests is to allow the agency to be notified formally of the contractor's readiness for inspection. These tests should indicate the performance of the system and may be conducted on prototype equipment. A more onerous test is the design qualification test which is a validation of the system design which will range from signal in space performance measurements to component rating measurements and analyses, electromagnetic compatibility measurements, and radiation exposure measurements. A major part of the design qualification is to test the system over its environmental service range.

After system qualification the agency will require a demonstration of the system reliability, which may be conducted at the contractor's facilities or in a controlled operational environment (eg agency designated airports). As this demonstration, usually to U.S. military standards⁷, takes a considerable amount of time - normally up to 12 months - the contractor may concurrently qualify other systems in anticipation of a favorable result. These subsequent qualification tests (sometimes referred to as type tests) consist of a subset of the full qualification tests.

MLS Service Life Cycle Testing

The service life cycle begins with the production testing of each individual system, and continues through the installation, commissioning and full periodic testing of each system. A commissioned system is one which has been tested from the bottom-up to meet the system, technical, and operational requirements. This part of the paper evaluates the testing necessary to take the MLS from the qualified design to certified operation: from system testing to operational testing. For the system service life cycle the testing requirements are reviewed for compliance with agency requirements and the ICAO Manual on MLS Testing⁸. This manual cites three categories of tests which can be applied to the requirements for testing in the service life cycle: factory, ground, and flight testing. The system characteristics which are tested in each of these categories are summarized briefly in Table 2. The allocation of tests to the categories shown represents the minimum testing necessary to achieve the desired operational category of service. As the ground tests and flight tests are a verification of system performance and are not dependent on the system design they can be specified by the agency at the outset in the master test plan. This is important as the agency should be aware of the extent and scope of these tests, particularly if turnkey installation is required as part of the procurement.

Ground Testing

The ICAO documentation recommends that MLS alignment be performed primarily through ground measurements, which are achievable as a result of the high frequency of MLS and the higher positioning accuracy attainable by ground measurement. Additionally it is more practical to measure absolute power density on the ground, as calibration of the ground receiver system is less complicated, and to use

these measurements to predict the power density at the limits of coverage. Several European agencies envisage using mobile MLS test vehicles to perform all testing, leaving the aircraft to verify that the system supports its promulgated operational procedures. However, as the capability to perform alignment includes a receiver system and an independent positioning system to measure errors, some agencies find it more cost effective to use aircraft measurement techniques for alignment. Geographically larger countries like the U.S. and Canada have found it more cost effective initially to perform the majority of testing by aircraft measurements, because of the distance between airfields, although the U.S. will employ ground testing as MLS ground equipment implementation progresses.

Table 2 Summary of MLS Service Life Cycle Testing

<u>Test Category</u>	<u>Characteristics To be Tested</u>
•Factory Production	•Frequency •Spectrum •Function Sequence •Function Timing •DME Reply Efficiency •DPSK Modulation •Midscan Timing •Scanning Beam Shape •DME Pulse Shape •DME Pulse Spacing •DME CW & Echo Suppression •Residual Radiation •Precision Angle Accuracy •Precision Distance Accuracy •Effective Radiated Power •Sidelobes •Data Coding •Monitor and Control •Remote Maintenance Monitoring •Interfaces
•Ground	•Data Coding •Precision Angle Accuracy •Precision Distance Accuracy •Effective Radiated Power •Function Synchronization •DME Reply Efficiency •DME Pulse Spacing •DME CW & Echo Suppression •Executive Monitoring •Monitor Control Actions •Station Frequency •Station Identification •Remote Maintenance Monitoring
•Flight	•Angle Accuracy •Distance Accuracy •DME CW & Echo Suppression •Relative Power Levels •Clearance Regions •False Courses •Monitor Limits •Facility Identification

Flight Testing

Flight inspection is the final verification that a system meets the operational requirements. There are two types of MLS flight inspection; the first is to test the guidance throughout the MLS coverage volume - this flight effectively samples the entire coverage volume. The second type of MLS flight inspection is more intensive and is a verification of the guidance

quality throughout all published procedures. Flight inspection throughout the coverage volume is performed using partial orbits and crossover flights. The intensive testing of an approach procedure is similar to the current flight inspection techniques for testing ILS displacement sensitivity and course structure.

Recurring and Nonrecurring Tests

System commissioning consists of a full ground test and flight test. Also, site modeling techniques may be used to estimate the location of any regions of out of tolerance guidance for flight inspection. This site modeling may be part of, or use results of, site modelling performed as part of the initial planning for the site and equipment selection. There are three generic types of MLS recurring tests⁹. These tests are referred to as long term, periodic, and routine tests, and their actual recurrence will vary between agencies for logistic and operational reasons. For long term testing a typical periodicity is 12 months and would consist of a full ground test and a flight test. For periodic testing a typical periodicity is 3-4 months and would consist of a subset of the ground tests. A typical periodicity for routine testing is daily or weekly and these tests would not rely on independent field measurements but should make use of MLS built in test equipment. Each of these tests is summarized in Figure 2.

TEST	TECHNIQUE	MODELING	FLIGHT	GROUND	BUILT IN TEST EQUIPMENT
TYPE					
COMMISSIONING					
LONG TERM					
PERIODIC					
ROUTINE					

Figure 2 MLS Tests and Techniques

Conclusions

In this paper individual system requirements have been evaluated with reference to the operational need and operational requirements for the system. The relationship between the operational and system requirements should be used as a framework to plan for comprehensive testing. By planning from the top-down testing can be performed from the bottom-up to ensure that the system meets all of the requirements. For MLS, the majority of testing is performed at the factory and on the ground, after installation, but flight inspection is the final verification that the system is performing satisfactorily in its operational environment.

Top-Down Test Planning - Bottom-Up Testing

System testing must be considered by the procurement agency throughout the equipment design life cycle as well as the service life cycle. In evaluating the requirements for MLS testing, the top-down processes of specifying requirements and designing to requirements have been applied to system testing, and it is concluded that a top down approach should be taken for system test planning. This approach is to develop a master test plan which the agency should consider at the outset of the procurement program. If possible, the plan should not simply be left for the contractor to write and the agency to edit and approve. The procurement specification requirements should form the basis of the test plan, and therefore should be written when the procurement specification is written. In the procurement specification, requirements should be quantitative wherever possible in order to define the test standards and the test demonstration procedure or analysis technique. The master test plan should include all of the requirements for testing to take the system from qualified design to certified operation. The agency should be aware of the extent and scope of tests required for the service life cycle of the system, particularly if turnkey installation is required as part of the procurement. As ground and flight tests are a verification of system performance and are not dependent on the system design they can be specified by the agency at the outset in the master test plan.

References

1. Operational Requirements For a New Non-Visual Precision Approach and Landing Guidance System for International Civil Aviation, Report of the 7th Air Navigation Conference, WP-90, Agenda Item 3, Appendix A, Montreal, 1974.
2. International Standards, Recommended Practices and Procedures for Air Navigation Services, Aeronautical Telecommunications, Annex 10 to the Convention on International Civil Aviation, April 1985.
3. Military Standard, MIL-STD-490A, Specification Practices, Department of Defense, U.S.A., June 1985, pp. 6-26.
4. *ibid*, pp. 31-37.
5. Military Standard, MIL-STD-1521B, Technical Reviews and Audits for Systems, Equipments, and Computer Programs, Department of Defense, U.S.A., June 1985.
6. Federal Aviation Administration Specification FAA-G-2100d, Electronic Equipment, General Requirements, U.S. Department of Transportation, March 1989, pp. 73-81.
7. Military Standard, MIL-STD-781C, Reliability Design Qualification and Production Acceptance Tests: Exponential Distribution, Department of Defense, U.S.A., October 1977.
8. Draft Manual Material on MLS Testing, ICAO All Weather Operations Panel Twelfth Meeting, WP/516, Montreal, November 1987.
9. A. E. Smith, "Review of All Weather Operations Panel Work on MLS Testing", in Proceedings of the Fifth International Colloquium on Flight Inspection of Navigation Aids, Paris, May 24-27, 1988, pp. 119-126.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

THE EFFECT OF ERROR TOLERANCES ON MLS ARRAY ANTENNA

M.Z. El-Gamal and C. Bowman

Microwave Landing System Project, Transport Canada
110 O'Connor Street, Ottawa Ontario, Canada K1A 0N8

Abstract

For Microwave Landing System (MLS) antennas, the analysis of the performance must take into account the effect of manufacturing errors inherent in the system as well as the effect of random failures of components in the radiating structure. In this work, the effect of element positioning, amplitude, and phase errors as well as the effect of failed elements on the pointing angle, sidelobe levels and beam width will be presented and analyzed for an MLS linear array antenna. The antenna considered has 56 elements with a symmetric amplitude distribution about its centre and the phase gradient across the array is assumed to be generated using digital four bit phase shifters. Random errors, within various tolerances, are added to the nominal position, amplitude and phase of each array element.

The pointing errors are found to be dependent on the phase errors and the results manifest a sinusoidal relationship between the element position and the pointing error. The amplitude of this error is dependent on the beam pointing angle and it is additive in nature. The amplitude tolerance and single element failures at zero degree pointing angle have no significant effect on the pointing error. However, the effect is greater at other angles considered.

The results show that the effect of the amplitude and phase errors on the beam width is negligible. However, element failures near the antenna centre narrow the beam width and element failures near the edge increase it. Sidelobe levels are increased by 1 to 3 dB over a 1 dB amplitude tolerance and by 3 to 4 dB over a 10 degree phase tolerance range and hence still within acceptable limits. The effect of the number of failed elements in the performance of the antenna array will also be presented and analyzed.

In addition, an empirical formula is developed to predict the pointing errors caused by element failures. The predicted pointing errors are modeled and found to be in a good agreement with the simulated errors.

1. Introduction

The Microwave Landing System (MLS) has been demonstrated to be a highly precise guidance system and it is expected to replace the Instrument Landing System (ILS) by the year 1998 as a precision landing approach to aircrafts in major airports worldwide. Errors, however, caused by equipment imperfections and propagation effects are introduced in the course of generating and detecting the MLS guidance information. These errors are essential for an antenna design and influence the construction tolerances which determine the overall performance and cost of the MLS antenna.

The effect of random errors on the performance of an antenna array has been studied first by Ruze [1] employing a statistical approach applied to a large number of identical antennas. Ruze's work has been extended by Schanda [2] to treat both amplitude and phase errors. Bao [3] has also extended Ruze's work and obtained expressions to study the effect of phase errors on the average power pattern taking into account the influence of illumination taper and correlation interval on the radiation

characteristic. His results show that for a larger correlation interval, the effect of errors becomes less critical. Smith and Guo [4] presented a comparison of different methods for randomizing phase quantization errors. They found that the so called "phase-added method", where a random start phase is introduced for each element in the feed section, gave the best performance but required known start phases at each element. Kelly [5] summarized and analyzed various MLS error source relations in terms of its fundamental system parameters. He also discussed the guidance accuracy considerations for the Microwave Landing Systems.

The ideal MLS antenna is characterized by an aperture having exact amplitude and phase distributions. In practice no antenna, however, is free from errors and these errors, regardless its cause, may affect the amplitudes and phases across the array. Amplitude and phase errors of a signal across a phased array antenna have effects on power, width, shape and pointing direction of the main beam as well as on the sidelobe level.

The purpose of this paper is to investigate the effects of element positioning, amplitude and phase errors as well as the failed antenna elements on the performance of an MLS azimuth antenna array in terms of its pointing angle, sidelobe level, and beam width.

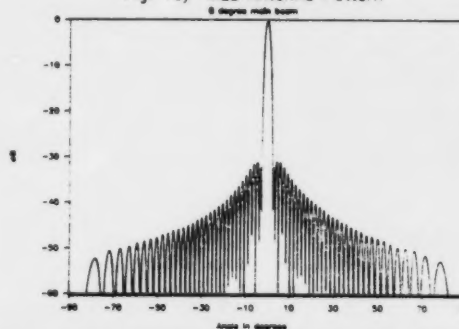
2. MLS Antenna Model

For an array of N radiating elements, the far field $E(\theta)$ is given by:

$$E(\theta) = E_0(\theta) \cdot \sum_{i=1}^N a_i e^{j\psi_i} e^{jkx_i \sin\theta}$$

where $k=2\pi/\lambda$, a_i is the amplitude of the excitation coefficient of the i^{th} element, x_i is the coordinate of the i^{th} element, ψ_i is the phase of the excitation coefficient of the i^{th} element, and $E_0(\theta)$ is the element pattern in θ direction. MLS azimuth antenna considered in this study has $N=56$ elements. The interelement spacing d is 0.593λ at 5.06 GHz and the scan angle is θ_0 . The corresponding half-power beam width (HPBW) is 2.03° . A software model written in Pascal and run on a VAX-11/780 Computer, is developed for this antenna array. The antenna radiation pattern is calculated in the plane of the scanning beam, i.e., the horizontal plane for the azimuth antenna. Figure 1 (a,b) shows the obtained radiation pattern at 0° and 40° scan angles respectively corresponding to four bit digital phase shifters.

Fig. 1(a) MLS Antenna Pattern



The developed program produces data files that specify the physical position, the amplitude, and phase angles of each element of the array for each position the main beam is to be pointed at. Random errors within some tolerances, can be added to each element's position, amplitude, and phase in order to

pointing errors when the beam pointing angle is 0° . The range of these errors for a single element failure falls within a range of $\pm 0.007^\circ$. Element failures, however, raise the sidelobe levels. This is due to the fact that when an antenna element does not have a corresponding radiating element on the opposite side of the array, a destructive interference will be of a lesser degree than that for the case of a uniform array and hence higher sidelobe level results in. The effect on the sidelobe level is more significant when elements near the array centre fail due to the Taylor series power distribution. It is also to be noted that element failures near the antenna array centre narrow the beam width and failures near the array edges increase it.

Comparing the element failure induced pointing error curves to the phase data shows that the pointing errors increase to a peak and then change sign at the element position where the phase increment between adjacent elements changes. At such a position, one expects the phase increment between elements changed to adjust the phase to the nearest quantized level. The pointing error $P_e(\theta, i)$ at a beam angle θ due to a failure of the i^{th} element can be expressed as:

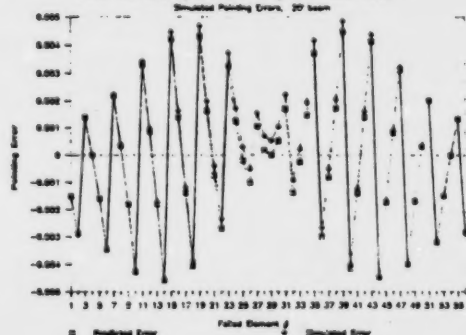
$$P_e(\theta, i) = K S(i) A(\theta) Q(i)$$

where K is a constant, $S(i)$ is the shape factor, $A(\theta)$ is the pointing angle factor at a beam angle θ , and $Q(i)$ is the phase quantization error on i^{th} element.

Curves such as those shown in Fig. 7 are normalized and used to determine $S(i)$, and are then divided by the corresponding data point values, with the curve for 0° beam as a reference. The average rates increase of each curve relative to the reference curve is then determined and taken as a measure of $A(\theta)$. With the 29^{th} element as a reference phase of 0° , the theoretical phase differences between adjacent elements for a beam pointing at angles θ are calculated and the differences between these phase states and the corresponding antenna four bit quantized phase data are determined to give $Q(i)$. The constant K can be determined by substituting the actual pointing error from the simulated results and employing the above equation.

The predicted (calculated from the last equation) pointing errors utilizing the given equation and using $K = -0.00049$ are shown in Fig. 10 as they are compared with the simulated error. This figure shows a good agreement between the predicted and simulated results.

FIG. 10 Comparison of predicted and simulated pointing errors, 20° beam

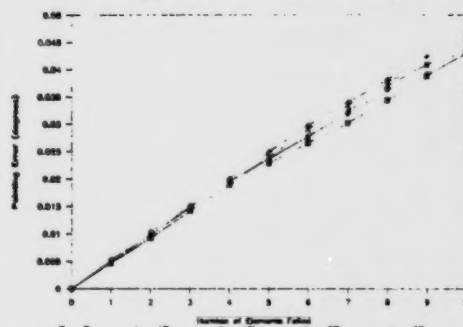


3.4.2 Multiple Element Failures: Multiple

element failures in groups of 2, 3, ..., 10 elements and the worst case results are reported in this study for pointing angles in the range $0^\circ - 40^\circ$. Fig. 11 shows the pointing error as a function of the number of failed elements and the results

suggest an approximate straight line relationship. This figure also suggests that the MLS antenna array can have multiple element failures, up to one third of its total number of elements, and still have acceptable performance.

FIG. 11 MULTIPLE ELEMENT FAILURES



4. Conclusions

A linear MLS antenna array was modelled and tolerance tests on positioning, amplitude and phase, and element failures were simulated to determine its effects on the pointing error, sidelobe levels and beam width. The pointing errors were found to be dependent on the phase errors and the results manifest a sinusoidal relationship between the pointing error and the element position. The amplitude of this error was found to be dependent on the beam pointing angle and can be considered additive. Moreover, the effect of the amplitude and phase errors on the beam width and sidelobe levels was found to be insignificant.

Element failures near the antenna centre narrow the beam width and those near the edges increase it. Sidelobe levels were increased by 1 to 3 dB over a 1 dB amplitude tolerance and by 3 to 4 dB over a 10 degree phase tolerance range and hence still within acceptable limits. The MLS antenna array considered can still perform adequately with a failure of its one third of its total number of elements.

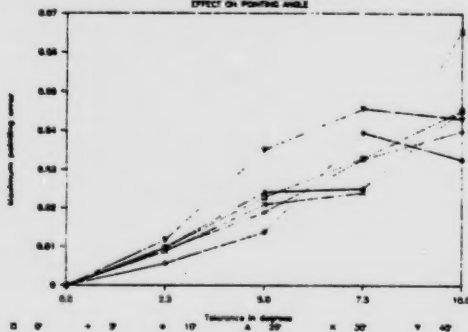
In addition, a formula was developed to predict the pointing error in terms of the pointing angle and phase quantization. The predicted and simulated results were in a good agreement.

References

- [1] J. Ruze, "Antenna Tolerance Theory: A review," Proc. IEEE, Vol. 54, pp. 633-640, 1966.
- [2] E. Schanda, "The Effect of Random Amplitude and Phase Errors of Continuous Apertures," IEEE Trans. Ant. Propg., pp. 471-473, 1967.
- [3] V.T. Bao, "Influence of Correlation Interval and Illumination Taper in Antenna Tolerance Theory," IEEE, Vol. 116, pp. 195-202, 1969.
- [4] Y.C. Guo and M.S. Smith, "Sidelobe Reduction for Phased Array Antennas Using Digital Phase Shifters, pt. I: One bit phase Weighting," Proc. IEE, Vol. 130, pp. 343-351, 1983.
- [5] R.J. Kelly, "Guidance Accuracy Considerations for the Microwave Landing System," J. Inst. of Navigation, Vol. 24, No. 3, 1977.

phase error distribution across the array to the pointing error produced, one can not the following points: a) pointing errors are minimum if the entire array have the same biased errors, b) larger pointing errors result when one side of the array has positive or negative biased errors, and c) the maximum pointing errors would occur when one side of the array has positive biased errors and the other side has a negative one.

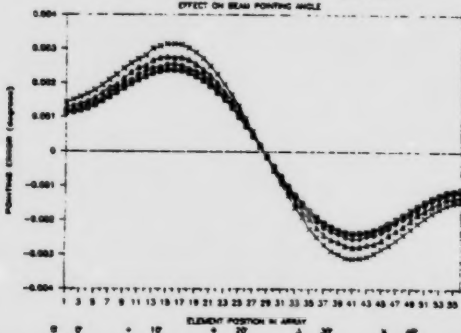
FIG. 6 ELEMENT PHASE TOLERANCES



For the phase tolerances used, that sidelobe levels (worst case) increased by 3 to 4 dB over a 10 degrees tolerance range and hence are still well within acceptable limits. The effect of the phase tolerances on the beam width is found to be negligible.

3.3.2 Phase Error Sensitivity vs Position in the Array: Five degrees are added to the nominal phase of one element in the array. The effect on the pointing error, sidelobe level, and beam width is then determined. This is repeated for each of the 56 elements in the MLS antenna considered to determine the relationship between the element position and performance degradation due to phase errors. This is done for 0°, 10°, 20°, 30°, and 40° beam pointing angles. It is evident from Fig. 7 that there is a

FIG. 7 PHASE ERROR VS. POSITION



sinusoidal relationship between the element position in the array and the pointing error that is caused by a phase error on that element. Elements near the centre or outside edge of the array are the least sensitive to phase errors. The magnitude of the pointing error is dependent on the beam pointing angle as it increases slightly with increasing beam pointing angles. Corresponding elements on opposite sides of the array centre induce symmetrical pointing errors. Figs. 8a and 8b show a negligible effect on the beam width respectively for 0° and 40° beams due to the variation of the phase error with the position of the array element.

FIG. 8a) PHASE ERROR VS. POSITION

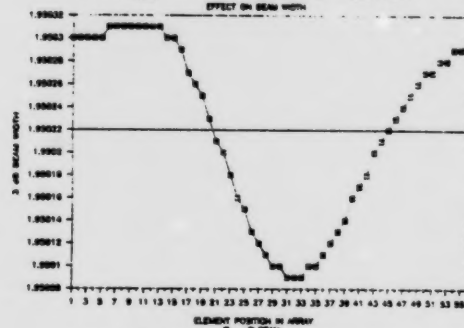
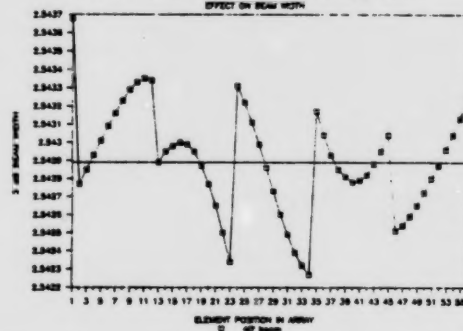


FIG. 8b) PHASE ERROR VS. POSITION



Additional tests suggest that the beam pointing errors induced by phase errors may be approximated as being additive. Addition, e.g., of 5° phase error on element number 3, 30, 49, and 50 causes pointing errors of 0.00119, -0.00044, -0.00171, and -0.00158 degrees respectively for a 0° beam angle. The sum of these errors is -0.00254°. When the same 5° phase error is added to the same four elements all at once, the pointing error is found to be -0.00256°.

3.4 Element Failure

3.4.1 Single Element Failure: An element failure is modelled by setting the amplitude of that element to zero. The effect on the pointing error, sidelobe level, and beam width is then determined. This is repeated for each of the 56 elements in the MLS antenna array. This is to determine the relationship between the element position in the array and the performance degradation due to element failure. This is again repeated for various beam pointing angles. Results illustrated in Fig. 9 show that element failures do not have a significant effect on the

FIG. 9 SINGLE ELEMENT FAILURES

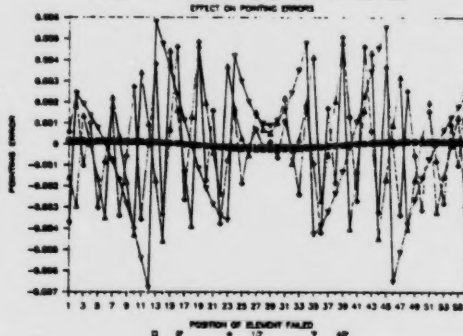
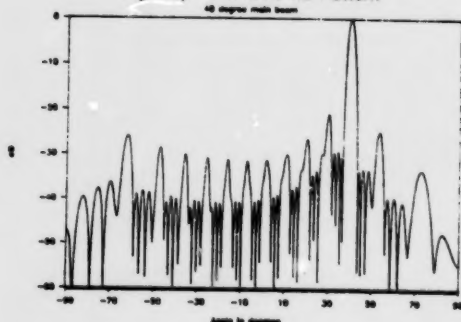


Fig. 1b) MLS Antenna Pattern



simulate tolerances, respectively, on the alignment of radiating elements on the antenna frame, amplitude and phase errors. The effect of random errors within each tolerance are conducted 50 times and the worst case results are reported in this work.

3. Results And Discussion

3.1 Element Positioning

The MLS antenna array can be expanded or contracted to simulate thermal expansion or contraction of the antenna frame. Random errors within the tolerances: ± 0.25 , ± 0.50 , ± 0.75 , and ± 1.00 mm are added to the nominal position of each antenna element. This is to simulate tolerances on the alignment of radiating elements on the antenna frame.

Figures 2, 3 and 4 illustrate the effect of the element positioning on the pointing angle, sidelobe levels, and beam width respectively. It is evident

FIG. 2 ELEMENT POSITIONING TOLERANCES

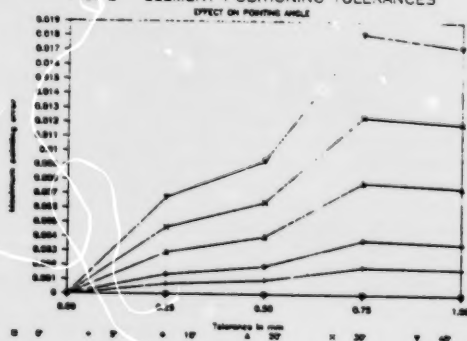


FIG. 3 ELEMENT POSITIONING TOLERANCES

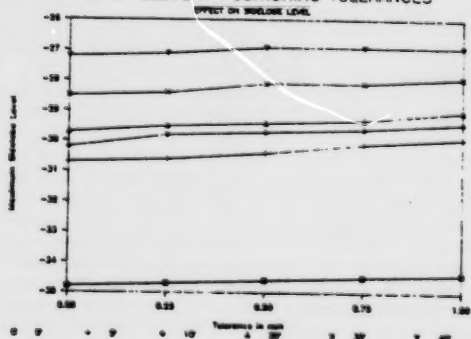
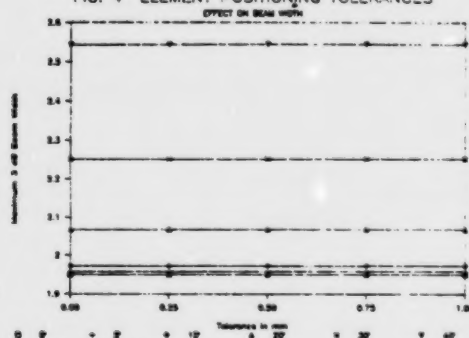


FIG. 4 ELEMENT POSITIONING TOLERANCES



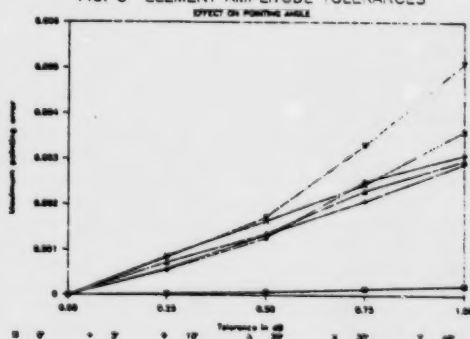
that the pointing errors increase as both the tolerance and pointing angles increase. It seems that an error in the element's lateral position produces an error in the phase path distance to the receiver which is proportional to the sine of the angle from centerline. Worst case sidelobe levels increased by less than 1 dB over a 1.0 mm positioning tolerance and hence is still within acceptable limits. The effect of element positioning on the beam width is found to be negligible as shown in Fig. 4.

3.2 Amplitude Tolerance

The program produces a data file which specifies the amplitude of each element of the array. It can calculate a Techebychev series amplitude distribution to meet sidelobe criteria. Alternatively, if a predesigned antenna is to be modelled, the amplitude (or phase) data can be read from a file or a keyboard. Random errors, within tolerances of ± 0.25 , ± 0.50 , ± 0.75 , and ± 1.0 dB, are added to the nominal amplitude of each antenna element to simulate tolerances in the corporate feed, RF cables, etc.

Fig. 5 illustrates that the amplitude error within the tolerances considered has a little effect on the pointing angle. Sidelobe levels, (worst case) increased by 1 to 3 dB over a 1 dB amplitude tolerance and hence are still within acceptable limits. The effect of the amplitude error on the beam width is found, again, to be negligible.

FIG. 5 ELEMENT AMPLITUDE TOLERANCES



3.3 Phase Errors

3.3.1 Phase Tolerances: Random errors are added to the nominal phase of each antenna element within tolerances of $\pm 2.5^\circ$, $\pm 5.0^\circ$, $\pm 7.5^\circ$, and $\pm 10^\circ$. This allows the simulation of tolerances in the net phase error due to sources such as phase shifters, RF cable length, machining of grooves in slotted waveguide radiation elements, etc.. Pointing angle errors are found to be dependent on phase errors as shown in Fig. 6. By comparing the

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

MLS SYSTEM ERROR MODEL IDENTIFICATION AND SYNTHESIS

R. J. Kelly
Allied-Signal Aerospace Company
Bendix Communications Division
Baltimore, Maryland

ABSTRACT

The All Weather Operations Panel (AWOP) of the International Civil Aviation Organization (ICAO) is currently studying the error source identification and modeling of the MLS signal. These studies serve two purposes. First, in order to certify aircraft for operations using the MLS guidance signals, regulatory agencies require documentation that the resulting performance will meet established requirements for the type of operations envisaged. Due to the high cost of flight testing, the bulk of this documentation is developed from simulation of the aircraft dynamics operating on inputs of the desired flight profile, wind gusts/shear, and navigational position-plus-error. In general, these simulations are highly detailed. They include system nonlinearities and (for certification purposes) are conducted under prescribed uniform input conditions. A critical element in these simulations is the MLS error source generator which is input to aircraft flight control system (AFCS) simulators during aircraft certification. The second purpose of these studies introduces a new perspective into navigation performance specifications—namely, error source identification. Introducing signal error source identification into performance specification would promote a better understanding not only of equipment design but the error mechanisms themselves. Armed with such information, the navigation equipment engineer could implement systems having performance characteristics closely approximating an optimum design. Error source model identification and generation is a tool applicable to ground and airborne equipment acceptance tests as well as flight inspection tests. Signal generation is the inverse operation of model identification. A methodology is described for identifying the MLS error components (both deterministic and random) from flight error records. After the model has been identified, the error components are synthesized to form an MLS error generator. The difficult step in the model development is identifying the random error components. A general technique is presented for developing a statistically equivalent MLS error generator using only a fraction of the data storage required to store the actual flight test data (400:1 data compression). This is an important consideration because the error generator must have in storage a library representative of MLS signals worldwide. The random error model is synthesized using the latest computer algorithms in System Identification Theory to develop Autoregressive Moving Average (ARMA) models. We also determine whether the random error model of flight data measured at a given runway is statistically equivalent to the signals generated at other airports. In other words, is there some invariant signal structure in the random error components which can be generalized to simplify the AFCS engineers design task. DME/P Flight test data recently collected by the Federal Aviation Administration (FAA) is analyzed to demonstrate the System Identification methodology. The technique applies to all navigation systems.

(1) INTRODUCTION

Two very different performance models are necessary to characterize aircraft approach and landing systems such as MLS. They are the Collision Risk Model (CRM) and the MLS Signal Model. In the former, the problem is to estimate the derived MLS

positional errors associated with very low probabilistic events (on the order of 10^{-7}). Obstacle clearance surfaces are based upon the quantification of these rare events. On the other hand, the MLS Signal Model needed for autopilot certification has an entirely different role. Its function is to estimate aircraft position errors which occur typically (i.e., with 95% probability) during aircraft approach and landing maneuvers. The subject of this paper is to model these typical error sources.

We describe three activities: (1) signal error source identification which permits data compression for efficient storage in a signal library; (2) signal generation or synthesis using the library of stored signals; and (3) multivariate hypothesis testing to determine the similarity or dissimilarity of signals at different airports. Signals from different airports (i.e., runways) having similar characteristics reduce the storage requirements of the signal library. Moreover, if certain signal characteristics persist across the ensemble of all airports, then the AFCS can be more closely matched to the MLS signal-in-space.

Signal model identification will make government regulatory agencies, equipment manufacturers and the user conscious of the mechanism by which error sources arise in navigation equipment designs. It is our experience that once we identify the error mechanism, practical means to control them close to their theoretical limits are always discovered. By exposing equipment performance to such analysis, the specifications themselves become consistent and operationally meaningful.

Although the theory and procedures discussed in this paper are not new and are well documented in the literature, they are new in the context of using them to define navigation system specifications. Evidence of the maturity of system identification theory is the availability of user friendly computer packages such as MatLAB[®] which includes the Identification Theory TOOLBOX[®] developed by Lennart Ljung⁴. The Precision Distance Measuring Equipment (DME/P) is used as the navigation example. System identification is simply to find the best fit of a preselected model given a set of noisy input and output data.

(2) OVERVIEW OF SIGNAL IDENTIFICATION
AND SIGNAL GENERATION

In this Section we present the critical steps used to analyze a flight error trace. The first set of operations determines the waveform model and is called model identification. The second set is an inverse operation which yields the waveform generator.

Figure 1 illustrates a waveform having both deterministic components and nonstationary random components. Figure 2 is a flow chart of the rigid identification and synthesis process. The deterministic components always dominate the data in the sense that their amplitudes are much longer than the amplitudes of the background random process. They can be removed by simple subtraction. If they are no larger than the background noise, they are indistinguishable from the random process. Deterministic error components include equipment bias, specular multipath error and garble effect errors. Although garble errors induced by DME pulse interference can be modeled as a Poisson random process, they are grouped with the deterministic components because they are easily removed by subtraction. Multipath errors are modeled by the RF interference of 2 DME pulses where the

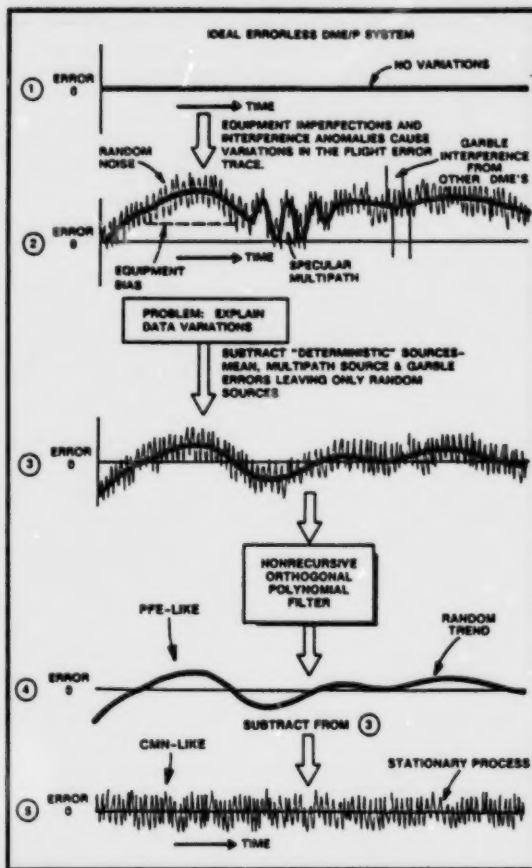


Figure 1. DME/P Signal Processing Task:
How to Model Stationary Random Sources

carrier frequency of the multipath pulse is doppler shifted with respect to the desired pulse. Only the interference starting time, desired signal amplitude, multipath delay time and the doppler frequency of the multipath need be stored in memory. Garble effects are isolated error spikes whose arrival times can be sprinkled randomly across the error trace using a Poisson distribution. Alternatively, their exact arrival times as measured from the original flight error trace can be stored in memory.

After the deterministic errors are subtracted, we are left with a nonstationary random waveform having two components. One is a low frequency process which is like a path following error (PFE).*

We call this time varying mean a random trend. The second component is a high frequency component which is like control motion noise (CMN).** We make the reasonable assumption that the CMN-like process is statistically independent with respect to the PFE-like process (see Kelly⁸).

* PFE—that portion of the guidance signal error which could cause aircraft displacement from the desired course and/or glide path. PFE frequency regions are 0-1.5 rads/s for DME/P.

**CMN—that portion of the guidance signal error which could affect aircraft attitude angle and cause control surface, wheel and column motion during coupled flight, but which does not cause aircraft displacement from the desired course and/or glide path. The frequency regions for CMN are 0.5-10 rads/s for DME/P.

The idea is to subtract the random trend from the raw nonstationary random waveform. The random trend is obtained by passing the raw random signal through nonrecursive orthogonal polynomial filter (Steffen⁹ or using the technique developed by Kitagawa¹⁰). What remains is a stationary waveform (the CMN-like component) which can be modeled by an Auto-Regressive Moving Average (ARMA) process. In other words, at this point we can use System Identification theory to determine the ARMA model coefficients. From this ARMA model a statistically equivalent waveform can be generated. Statistical equivalence means that the autocorrelation function of the simulated model output or its power spectrum are equal to the autocorrelation function of the input data or equivalently its power spectrum.

What do we do with the random trend component once it has been identified by the nonrecursive orthogonal polynomial filter? One cannot store the random trend at the sample rate f_s , because that would defeat the purpose of our task, i.e. to regenerate a statistically equivalent version of the random waveform using as few model parameters as possible. Storing all the waveform samples does not achieve this goal.

Three options are available:

1. Use data compression by sampling at $f_L \ll f_s$. Store the random trend samples having a compression ratio

$$K = \frac{f_s}{f_L}$$

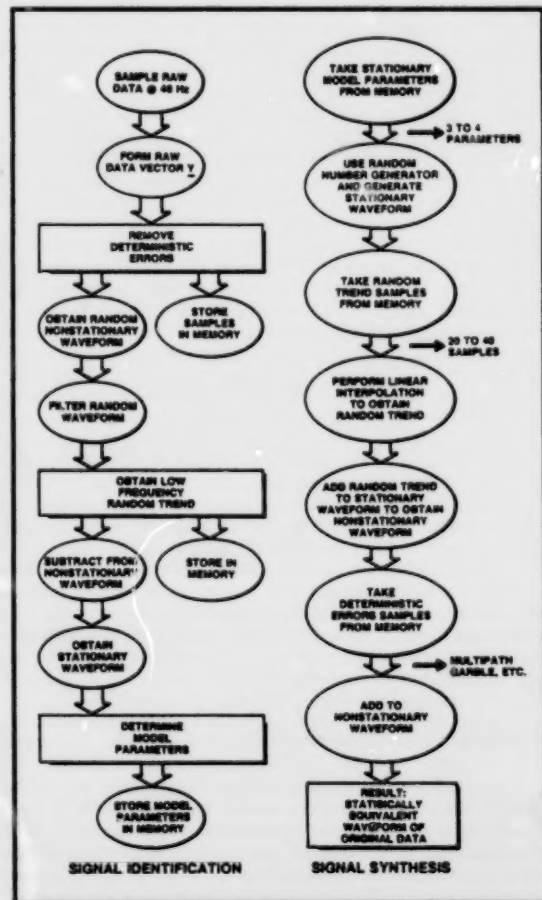


Figure 2. The Signal Modeling Process; Signal Synthesis is the Inverse of Signal Identification

as shown in Figure 2. One can do this because the low random trend is highly correlated, that is its information bandwidth is much smaller than f_c .

2. Perform two stationary ARMA analysis. Determine the ARMA parameters for the random trend as well as for the CMN-like random waveform (Kelly⁷).
3. Use the Box-Jenkins³ approach to time series identification which removes the nonstationary components by differentiating the raw random waveform. During the signal synthesis process, remove the differentiation operation by integrations.

An important point is that method 1 above will produce the random trend function exactly, while methods two and three can generate only equivalent realizations. On the other hand, method one requires that more parameters (data points) be stored than that required by methods 2 and 3).

(3) DME/P SIGNAL MODEL PROBLEM

The purpose of this Section is to present a brief example of what is meant by explaining the data and what is done with the unexplained data or residuals. System Identification Theory is the methodology which explains the data where the model of the system is the system error sources. The focal point is the linear model described in Ljung⁴ and Box³. In their concept the linear model is embodied in a filter transfer function which is computed from flight test data. The idea is that flight test data or receiver bench test data is the realization of a particular stochastic process which can be modeled by a linear filter. This means that when the filter is driven by a white noise signal it will generate data which is statistically equivalent to the original data. It is assumed that all of the deterministic sources have been removed beforehand and that the data is a single realization of a stationary process. Since this white noise input source has no signal structure of its own, the identification concept assumes that the entire signal structure is contained in the filter transfer function. As shown in Figure 3, the test data is run through a filter whose filter parameters are sequentially varied until the serial correlation in the data is removed. If this process is successful, the filter output should approximate a small white noise signal. This signal remnant is called the "residuals" in the statistical literature. When these residuals reach a minimum as determined by the FPE^{*} criteria and are "white noise like," the final settings of the filter coefficients define the optimum inverse filter of the input data. As mentioned above, one test of "whiteness" is to plot the autocorrelation function of the residuals and note whether 95% of its side lobes are contained within confidence limits which were determined from a pure white noise signal. Once this condition is satisfied, we can say the input data has been identified by a linear filter.

Having derived the optimum inverse filter $H^{-1}(z)$, the next step is to invert its transfer function. The resulting transfer function $H(z)$ is then driven by white noise and its output should be statistically equivalent to the original raw input data. This final step yields the MLS error signal generator which could be included in the aircraft flight control system simulations.

• Example DME/P System Identification Process

This Section illustrates the System Identification procedures for a DME/P flight test error record. The error data was taken from the E-System ICAO standardized preproduction ground transponder and the Bendix Avionics Pre-ICAO standardization interrogator on runway 31 at the FAA Technical Center,

*Optimum filter selection contains the Find Prediction Error (FPE) defined by Akaike¹

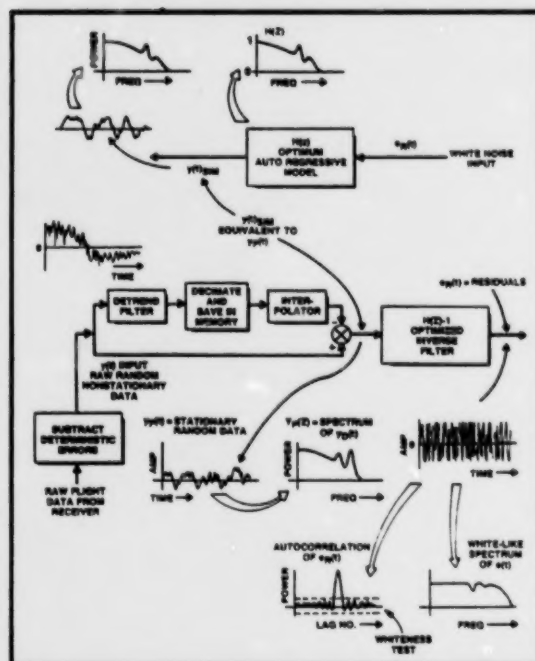


Figure 3. Statistical Identification/Synthesis Process

Atlantic City, New Jersey. Aircraft position reference was a laser tracker where DME/P error = (DME/P - Laser Tracker) in feet. The interrogator rate was 40 Hz for the Final Approach Mode (FA) and 16 Hz for the Initial Approach Mode (IA). Data sample replies were recorded at the interrogation rate. Three runs were recorded on November 11, 1988; all were 3° straight-in approaches with an aircraft speed of 240 to 250 ft/sec. Only the FA mode data for Run 1 will be analyzed. As shown in Figure 4, it extends almost to 5 nmi and has both deterministic errors as well as nonstationary random components. Error sources easily identified in the figure are ground and airborne equipment bias and noise errors (e.g. thermal noise), garble effects, diffuse multipath and low level multipath with low frequency fluctuations. Figure 4 is the FA mode errors remaining after the 110 ft. bias and garble errors were removed. Forty feet of the bias was due to the uncompensated antenna cable.

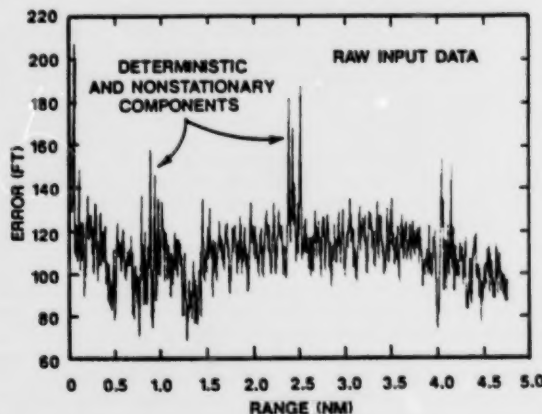


Figure 4. DME/P Raw FA Mode Data

The garble or outlier errors were removed, leaving a nonstationary random flight record. Garble/outlier "like" errors had to exceed ± 40 ft and had to be isolated "spikes" before they were removed. It is clear that the nonstationarity exhibited in Figure 5 is due to a small amplitude low frequency random trend. This random trend was removed by a reduced version of the nonrecursive orthogonal polynomial filter discussed earlier. Noise bandwidth was 1 rad/sec. Subtraction of the random detrend function yields the CMN-like random waveform shown in Figure 6. Only 36 samples were necessary to store the random trend waveform—this is a data compression of 120:1.

Model identification using the MatLAB® Identification TOOLBOX® resulted in an AR(3) model; the optimization process was the Modified Covariance method described in Kelly⁷. Its residuals and identification filter parameters are shown in Figure 7. Figure 8 illustrates how much of the FA mode random data was explained using input and output ACF's.

As shown, the ACF of the residuals passes the whiteness test. The equivalent description in the frequency domain is shown in Figure 9. If the AR(3) model of Figure 7 is accurate, its theoretical transfer function will closely approximate the CMN-like data spectrum. The result of the optimization process was an AR(6) filter with minimum FPE = 15.36—the 3 pole AR filter was selected because it best satisfied Box³ criteria of Parsimony, i.e., supply the smallest possible number of parameters for adequate representation. An AR(3) model yields an FPE = 15.39 and a loss function $\sigma_e^2 = 15.34$. The loss function is the variance of the residuals shown in Figure 7. Thus, Figure 4 raw data has been decomposed into 5 components, each requiring several parameters which are to be placed in storage. They are

COMPONENT	NUMBER OF PARAMETERS*
Bias	1
Garble	6
Random Trend	36
AR(3)	3
Variance of Residuals	1
Total	47
Number of FA Mode Run 1 Data Samples	4,320

*The comparison should be based upon the total bytes required.

Thus only 47 parameters have to be saved in order to generate a flight error trace which is statistically equivalent to the actual record of 4320. This comprises the signal identification step outlined in Figure 2.

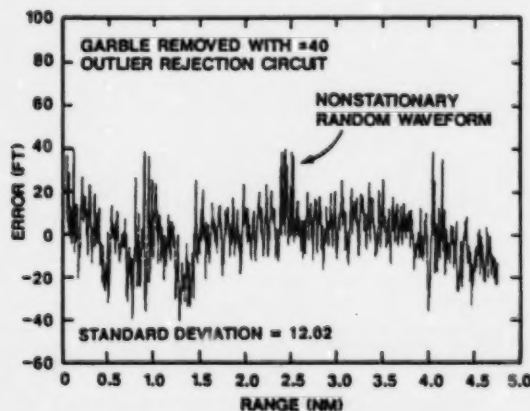


Figure 5. DME/P FA Mode With Garble Errors and Bias Removed from Figure 4

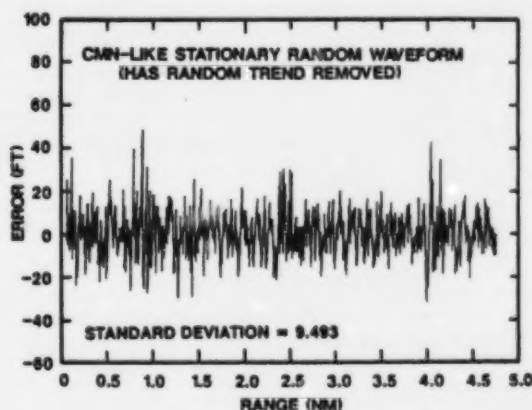
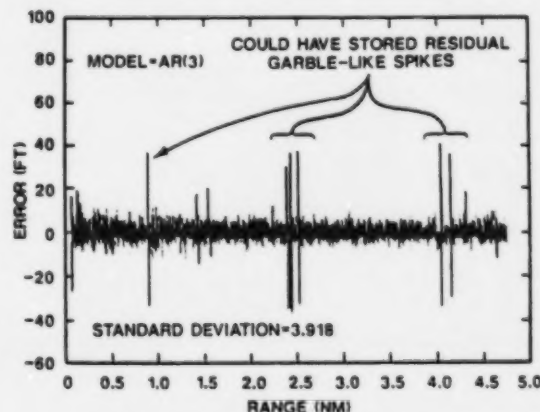


Figure 6. DME/P FA Mode Detrended Waveform



3-Pole AR Filter transfer function, $H(z)$, derived from data shown in Figure 6

$$H(z) = \frac{1}{1 - .94989z^{-1} + .0163z^{-2} + .0291z^{-3}}$$

Figure 7. DME/P FA Mode Residuals and 3-Pole Identification Filter

The signal synthesis steps are also outlined in Figure 2. The first step is to generate a statistically equivalent waveform of the CMN-like data using a random number generator. One such realization is shown in Figure 10 (compare with Figure 6).

Adding the random trend function to the Figure 10 simulated data, synthesizes the nonstationary waveform shown in Figure 11. Compare with Figure 5. When the bias and garble is added to Figure 11, the complete synthesized FA mode data has been reconstructed as shown in Figure 12. Compare with Figure 4.

Even less data has to be stored if the random trend PFE-like waveform is also modeled instead of reconstructing it exactly using a linear interpolation algorithm. This signal identification process is functionally described in Kelly⁷. However, the simulated nonstationary waveform will not necessarily be visually equivalent because the random trend sample path will not have the same realization as the original random trend. The signal storage would be 13 parameters rather than the 47 required earlier.

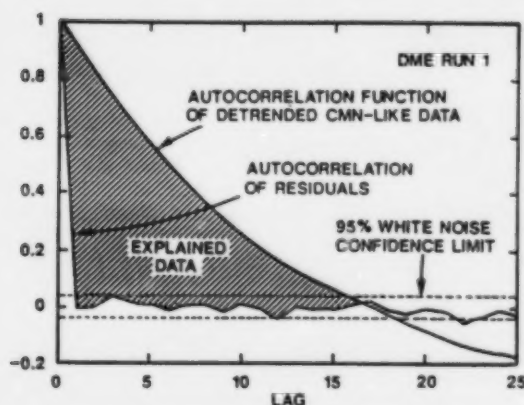


Figure 8. Comparison of Detrended CMN-Like Autocorrelation Function (ACF) With ACF of Residuals

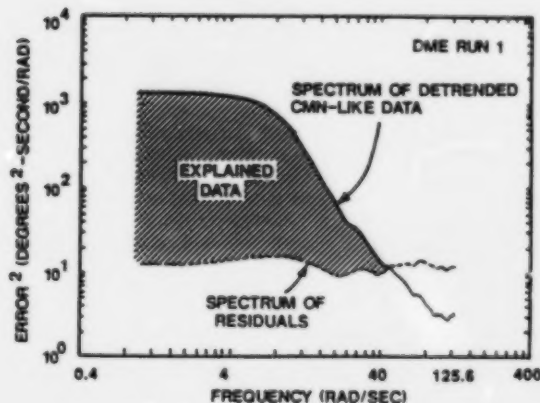


Figure 9. Comparison of Total CMN-Like Data Spectrum and Its Unexplained Residuals

The analysis of variance components (1-sigma) are

	Variance	Std. Deviation
1 Nonstationary Data	144.2 ft. ²	12.01 ft.
2 Random Trend	54.9 ft. ²	7.41 ft.
3 CMN-like	90.1 ft. ²	9.49 ft.
4 Residuals (Unexplained)	15.4 ft. ²	3.92 ft.
5 Explained	74.7 ft. ²	8.64 ft.

The mean of 1, 2, and 3 = 0. Note subtracting 2 from 1 yields 89.1 ft.² which is approximately equal to the CMN-like errors of 90.1 ft.². The signal identification process above has "tagged" all the principle error source and statistically quantified their magnitudes. Comparing Figure 12 with Figure 4 shows that the error signal has been successfully synthesized. The ICAO Annex 10 SARPS international standards specifies that the PFE ≤ 100 ft. (2-sigma) and the CMN ≤ 60 ft. (2-sigma). As shown above, the 2-sigma errors are well within these standards. PFE = 70 ft. bias + 14.8 ft. (random trend) = 84.8 ft. CMN = 18.98 ft. (CMN-like data). The garble errors ($> \pm 40$ ft.) increased the CMN error from 9.49 ft. to 11.8 ft.

(4) DME/P SARPS SYSTEM

This section analyzes a DME/P system for which both the ground and airborne equipment satisfied the ICAO Standards and Recommended Practices and Procedures (SARPS), the equipment being the E-Systems ground transponder and the SEL (Federal Republic of Germany) interrogator. Figure 13 is the flight error trace recorded at the Lebanon, NH airport. Flight tests using the ICAO SARPS approved system were also recorded in Atlantic City. Their CMN-like AR(3) synthesized filter parameters from the flight test data are:

	a0	a1	a2	a3
Lebanon, NH	1.0	-0.6298	-0.0621	-0.376
Atlantic City, NJ	1.0	-0.6204	-0.0513	-0.351

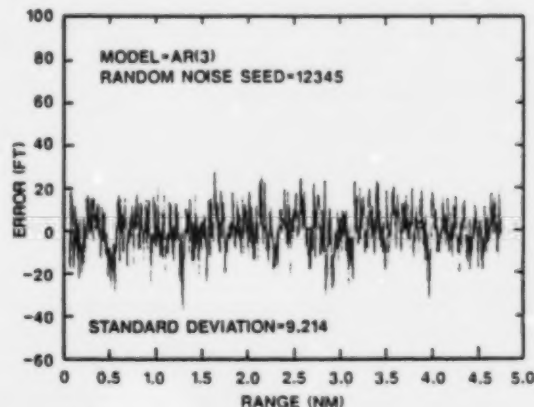


Figure 10. Detrended CMN-Like Simulated Data

(5) GROUND TRANSPONDER ERROR MODEL

Figure 14 shows raw data bench tests taken on the E-systems SARPS compliant ground transponder. Using this data, a transponder error model filter was identified. The equipment was well within the SARPS PFE and CMN transponder delay accuracy standards. Figure 15 shows the ACF of the residuals after the filter was modeled and the ACF of the model input data. Note only very small differences exist between these autocorrelation functions.

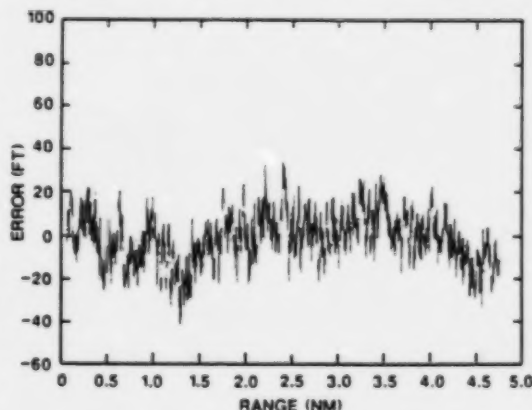


Figure 11. CMN-Like Simulated Data of Figure 10 With Random Trend Added to Give Nonstationary Random Waveform Compare with Figure 5

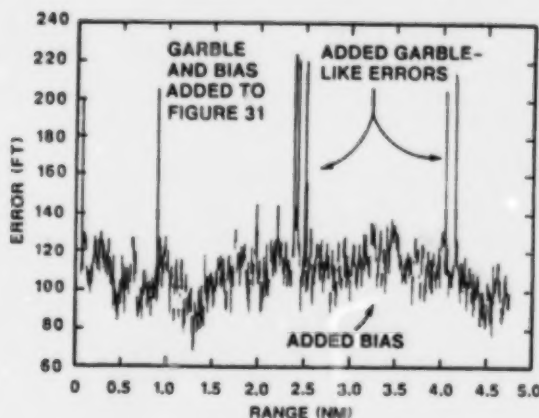


Figure 12. Final Simulated Waveform of Original FA Mode Data (Figure 4) With Bias and Garble Added

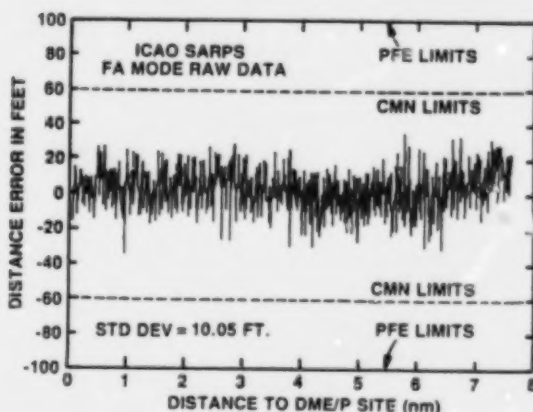


Figure 13. ICAO SARPS System FA Mode Raw Data—Lebanon, NH

Since the autocorrelation side lobes of the model input data are within the 95% "white noise" autocorrelation side-lobe limits, the E-Systems raw transponder errors are uncorrelated. This conclusion supports the design error analysis assumption. It is an example of how System Identification Theory can extend the performance analysis specification of ground navigation equipment beyond the simple quantification of performance in terms of bias and variance. Note the residuals cannot be partitioned further.

Once the data is explained to the extent that the unexplained residuals are statistically independent (white noise) nothing more can be done to the data. Even though several statistically independent error sources (e.g., thermal noise and quantization effects) make up the unexplained data, they cannot be partitioned or identified. To see this, assume that the variance of the residuals σ_k^2 is equal to $a\sigma_1^2 + b\sigma_2^2$ where σ_1^2 and σ_2^2 are the variances of the thermal noise and the quantization effects, respectively, and are unknown. Perform an analysis on two data samples. After regression analysis, there results two variances of the residuals σ_{k1}^2 and σ_{k2}^2 . Then,

$$\begin{aligned}\sigma_{k1}^2 &= a\sigma_1^2 + b\sigma_2^2 \\ \sigma_{k2}^2 &= a\sigma_1^2 + b\sigma_2^2\end{aligned}$$

This equation cannot be solved to obtain σ_1^2 and σ_2^2 . Additional information is necessary such as an independent measurement of σ_1^2 or σ_2^2 . The systematic and more accurate approach is to change the condition of the experiment for the second sample (e.g., increase the temperature). This, however, is precisely what the Analysis of Variance (ANOVA) is all about. In any event, it is not necessary to partition the residuals when developing the signal model.

(6) COMPARISON OF ICAO SARPS SYSTEM MODELS FROM TWO DIFFERENT SITES

The Section 4 ICAO SARPS System filter designs were developed from two different samples of data originating from two different airports—Lebanon, NH, and Atlantic City, NJ. The Hotelling Generalized T^2 test uses data from these runs to test if the filter designs are, in fact, samples from the same population. We call this the null hypothesis which states that the differences between the two sets of filter coefficients are assumed to be zero. The differences between the estimated filter coefficients are due to the errors arising from a finite sample size. A test statistic, T^2 , is calculated from the data and compared with a critical value T_{α}^2 . If $T^2 < T_{\alpha}^2$ then the null hypothesis is accepted. If $T^2 \geq T_{\alpha}^2$ the null hypothesis is rejected.

In this problem neither the means $\bar{Y}_1$ and $\bar{Y}_2$ nor their covariance matrices are known; they must be estimated from the data.

Let $\hat{Y}_1 = [-0.6298, -0.0621, -0.3760]$ represent the Lebanon, NH filter given in Section 4 and let $\hat{Y}_2 = [-0.6204, -0.0513, -0.0351]$ represent the Atlantic City filter given also in Section 4. Then the T^2 statistic is

$$T^2 = (\hat{Y}_1 - \hat{Y}_2)^T S^{-1} (\hat{Y}_1 - \hat{Y}_2) \quad \text{where} \quad (1)$$

$$S = \left(\frac{1}{N_1 + N_2 - 2} \right) [(N_1 - p)A_1 + (N_2 - p)A_2] \quad (2)$$

and $p = 3$ is the number of filter parameters (Anderson⁶). The matrix

$$A_1 = \begin{bmatrix} \hat{y}(0) & \hat{y}(1) & \hat{y}(2) \\ \hat{y}_1(1) & \hat{y}_1(0) & \hat{y}_1(1) \\ \hat{y}_1(2) & \hat{y}_1(2) & \hat{y}_1(0) \end{bmatrix} \quad (3)$$

is the sample covariance matrix for the Lebanon, NH flight test data and A_2 is the sample covariance matrix for the Atlantic City data. $\hat{y}(k)$ is the sample autocovariance function for the k th offset. The connection between the filter coefficients and the flight test

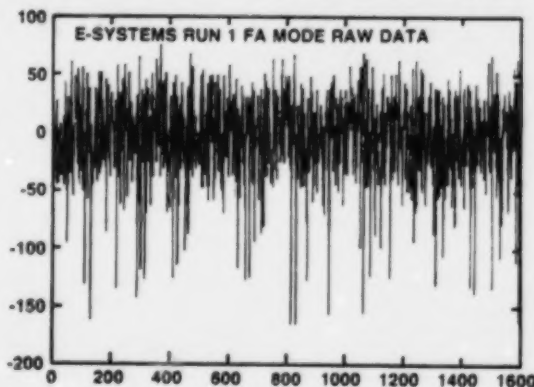


Figure 14. E-Systems Ground Equipment Bench Tests Raw data

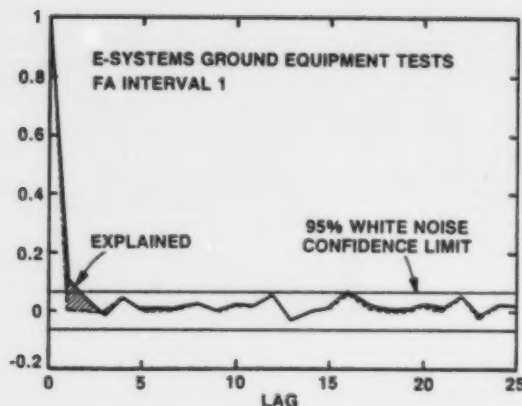


Figure 15. Comparison of ACF of Raw Data with the ACF of the Residuals for E-Systems Ground Test

data is established in Fuller⁵, page 335. For $N_1 = 2280$ and $N_2 = 1380$ and

$$A_1 = 1.8 \times 10^5 \begin{bmatrix} 1.000 & 0.65 & 0.44 \\ 0.65 & 1.00 & 0.65 \\ 0.44 & 0.65 & 1.00 \end{bmatrix}; \quad (4)$$

$$A_2 = 1.9 \times 10^5 \begin{bmatrix} 1.00 & 0.67 & 0.49 \\ 0.67 & 1.00 & 0.67 \\ 0.49 & 0.67 & 1.00 \end{bmatrix}$$

Eq. 2 becomes $T^2 = 0.0014$. This value is compared with

$$T_0^2 = \frac{(N_1 + N_2 - 2)p}{N_1 + N_2 - p - 1} F_{p, N_1 + N_2 - p - 1}(\alpha) \quad (5)$$

where $F_{p, N_1 + N_2 - p - 1}(\alpha)$ is the F-distribution and α is the significance level, the critical region is $T^2 \geq T_0^2$.

The critical $\alpha = 5\%$ value for this data was $T_0^2 = 8.28$. Since $T^2 = 0.0014$ does not exceed the critical value $T_0^2 = 8.28$, the null hypothesis is supported, implying that the CMN-like sources at the airports were similar. These results are not surprising because, as shown in Section 4, the filter coefficients for each run are similar, as are their sample covariance matrices A_1 and A_2 .

A Multivariate Analysis of Variance hypothesis test must be used when three or more filters are to be tested. The test statistic is the Wilk's likelihood ratio criterion. When only two filters are to be compared, the Wilk's test reduces to the Hotelling T^2 test.

CONCLUSIONS

A technique has been described which can in principle identify the error sources in navigation systems and can also perform the inverse operation of synthesizing a error generator whose output is statistically equivalent to the original error data.

ACKNOWLEDGEMENT

The author wishes to thank Janice Winter of Bendix Communications Division for performing the computer simulations and plotting the data. He also wishes to thank Dr. N. K. Nagraj and Dr. T. Mathews of the Mathematics and Statistics Department of the University of Maryland Baltimore County for pointing out the connection between AR filter coefficients and generalized hypothesis testing.

REFERENCES

- (1) Akaike, H. (1970), "Statistical Predictor Identification," Ann. Inst. Stat. Math, Vol. 22, pp. 203-217.
- (2) Akaike, H. (Dec. 1974), "A New Look at the Statistical Model Identification, IEEE Trans. Autom. Control, Vol. AC19, pp. 716-23.
- (3) Box, G.E.P. and Jenkins, G. M. (1970), *Time-Series Analysis, Forecasting and Control*. Holden-Day (McGraw-Hill Book Co., New York and Maidenhead, Eng.).
- (4) Ljung, L. (1987), *Systems Identification—Theory For The User*, Prentice-Hall.
- (5) Fuller, W., *Introduction to Statistical Time Series*, John Wiley, 1976.
- (6) Anderson, T., *An Introduction to Multivariate Analysis*, 2nd Edition, John Wiley, 1984.
- (7) Kelly, R. "MLS System Error Identification and MLS Signal Models," Ohio University Avionics Center Seminar, 3 Parts, January/March/April 1989.
- (8) Kelly, R., Cusick, D., "Distance Measuring Equipment and its Evolving Role in Aviation," *Advances in Electronics and Electron Physics*, Vol. 68, Academic Press, 1986.
- (9) Steffen, P., "On Digital Smoothing Filters: A Brief Review of Closed Form Solutions and Two New Filter Approaches," *Circuits System Signal Process*, Vol. 5, No. 2, 1986.
- (10) Kitagawa, G., Gersch, "A Smoothing Priors Time Varying AR Coefficient Modeling of Non Stationary Covariance Time Series," *IEEE Trans. Auto Control*, AC-30, No. 1, January 1985.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

FACTORY ACCEPTANCE TEST RANGE

Robert Janes, Malcolm MacNeil
Micronav Limited
Sydney, Nova Scotia, Canada

Abstract

In testing large antenna systems, including Microwave Landing Systems, the antenna range is often removed from the factory. This causes logistics problems in moving, packing and unpacking antennas. This paper describes the antenna range that Micronav Limited has developed, essentially right on the factory floor, to test systems as soon as they are assembled. The site selection criteria are reviewed along with the operational aspects of the range. A range designed for factory acceptance testing provides a cost effective alternative to duplicating a design approval range.

Introduction

In the manufacture of systems incorporating large complex antennas, such as the phased array antennas employed in Microwave Landing Systems, it is necessary to have an antenna test range where the system can be assembled and operated during the Factory Acceptance Tests. Where a company with an existing facility undertakes antenna work, it is often difficult to set up this sort of range adjacent to the existing factory area, thus leading either to a range situated well away from the factory or an elaborate range configuration such as roof top to roof top. Where a facility is being set up for antenna manufacture, one of the siting criteria can be access to land suitable for such a range. Through attention to range access and range operating procedures, an economy of logistics can be obtained to improve manufacturing efficiency.

If the range is intended for factory acceptance testing rather than design approval tests, its goal is to operate the equipment in its normal orientation and configuration and have the range measure the signals at the key points: at the limits of coverage and near the intended centerline and glide path angle. By designing a range adequate to test system functionality rather than trying to duplicate a design approval capability, there are also savings in the cost of the range.

This paper outlines Micronav's factory antenna range which has been designed to these principles and is currently being implemented at Micronav's facility in the Sydport Industrial Park, Sydney, Nova Scotia.

Site Selection and Preparation

Access to enough land for an antenna range immediately adjacent to the building was a key criteria for the selection of the location of Micronav's new plant after purchase of the company by Leigh Instruments in January of 1988. The other major criteria was to find a building suitable for Microwave Landing System production where the company could get set up in short order. The leading contender was a general purpose manufacturing building still at an early stage of construction which was located at the edge of the developed portion of the park. It overlooked well over a

thousand feet of land gently sloping away from the building which was undeveloped except for mounds of earth that had been placed there during the second World War to protect ammunition magazines. The magazines had previously been removed and investigation of the mounds proved that they were piles of soil which could simply be plowed down and spread over the range area. The old access road to the magazines provided a base for access roads to the range.

This site was selected as construction of the building was still at an early enough stage that the facility could be modified to a purpose built MLS manufacturing facility. The landlord was also agreeable to Micronav holding option of first refusal on the range and adjacent properties to guard against encroachment by future development of the park. Range preparation work began in the fall of 1988 on the first phase of the range. The range was first used in support of engineering testing in February 1989 and is expected to be fully implemented by early 1990. This phased approach allows the range to evolve to meet test requirements as the MLS equipment is developed.

Range Layout and Features

A plan view of the range is shown in Figure 1 and it can be seen that the range breaks down to three component parts:

- antenna deck
- field monitors
- range antennas

A fourth element, not shown on the layout, is a test area just inside the building adjacent to the antenna deck, where the range can be controlled and measurements gathered and processed.

Antenna Deck

The antenna deck is a concrete platform at the factory floor level at the rear of the building as shown in Figure 2. Access is provided by an overhead roll door that is sufficiently large to allow a completely assembled Azimuth or Elevation antenna mounted on its factory dolly to be rolled out onto the deck. The dolly is outfitted with bolt holes that align with threaded inserts in the concrete deck to insure consistent antenna placement for testing. The platform is large enough to have both an Azimuth and Elevation subsystem under test simultaneously. Individual power lines are brought to the deck to power the two subsystems independently.

The deck locates the equipment under test more than five feet above grade and fifteen feet away from the building. Both factors decrease the likelihood of multipath effects.

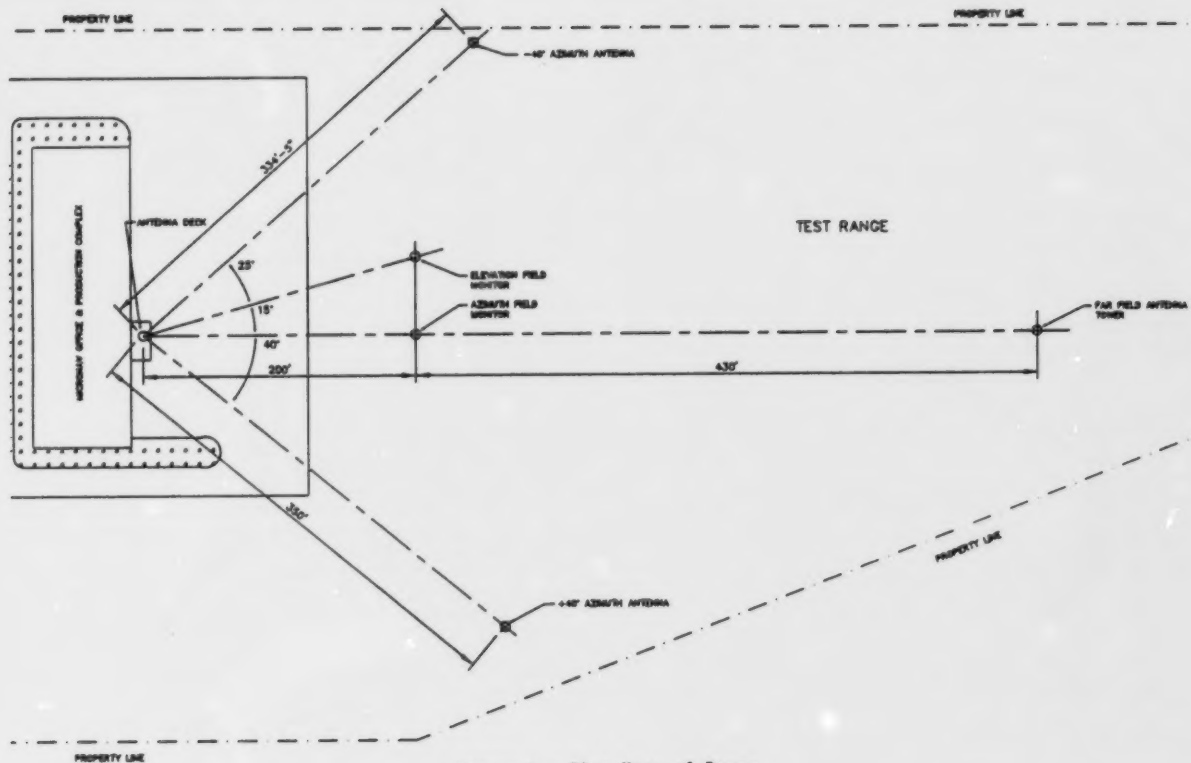


Figure 1 - Plan View of Range

Field Monitors

The field monitors are part of the deliverable equipment and must be included in the factory acceptance test. Two field monitor towers are located on the range two hundred feet out from the deck. The azimuth field monitor is located on the azimuth antenna centerline at an elevation angle of one degree so as not to shadow far-field measurements. The elevation field monitor is situated at the nominal glide path angle of three degrees but offset 15° in azimuth, again to avoid shadowing.

The towers are free standing (no guy wires) and are hinged at the bottom to allow them to be winched up and down to access field monitor modules. The field monitors are connected back to their respective subsystems through an underground conduit.

Range Antennas

The major range antennas are located at the far field antenna tower which is situated 630 feet from the deck which is a compromise between the two common expressions [1] for the start of the far field region for the elevation antenna:

$$R = D^2 / \lambda$$

and

$$R = 2 D^2 / \lambda$$

where

R = far field boundary range

D = size of the aperture

λ = wavelength

A 120 foot tower at this range provides for measurements up to 7.5° elevation. This tower will be outfitted with a track system to move the horn antenna vertically for elevation measurements. A mechanism for fine movement horizontally about the centerline is yet to be determined. Two additional receiving horns will be placed at the extremes of azimuth coverage (+40° and -40°) at an elevation of 3°.

The RF signals from all the range antennas will be brought to a switch feeding the MLS receiver in an enclosure at the base of the tower. Control of the switch and the receiver will be remote from the main building. Output from the receiver will be sent back to the main building for recording and analysis.

General Access

The range area itself is essentially flat ground so that measurements can be made at other azimuth angles by setting up temporary towers or by driving a suitably equipped vehicle across the range.

Range Accuracy

In order for the range to meet its goals of ease of use while providing sufficient accuracy to meet equipment testing requirements, the key element is maintaining repeatability of measurement. The design approach of fixing the transmitter under test as well as using permanent antenna towers meets this requirement.



Figure 2 - Elevation Equipment on Antenna Deck

Equipment Placement

The dollies that the equipment are mounted on utilize holes that align the equipment with threaded inserts in the Antenna Deck to obtain a repeatability that is well within the electronic boresighting capabilities of the equipment. As the equipment installation on the antenna deck simulates an actual field installation, it is necessary to go through the normal installation procedures of levelling and boresighting.

A unit whose performance has been measured on a high precision range can be used as reference to identify any anomalies on the factory range and indicate what improvements need to be made to remove them. This process will be carried out with equipment currently undergoing design approval tests when they are complete.

Antenna Towers

The towers used to support the test antennas are specified so that their mechanical characteristics will not contribute more than 0.01 degrees uncertainty to range measurements. This leads to specifications limiting movements of antenna structures to fractions of an inch. For this reason, the main far-field antenna tower will be guyed with the guys spaced to give maximum clearance to the receiving antennas on the tower.

Antenna Surveying

The range will be initially surveyed for the angles from reference to the antennas to an accuracy of 0.003 degrees using a theodolite. Measurements for moveable antennas will be made at regular increments and marked so that antennas can be reliably returned to these positions. Angular measurements will be re-surveyed every month if in regular use or at the start of a series of tests.

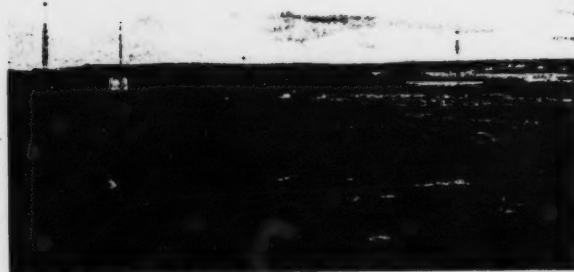


Figure 3 - Range Antennas

Summary

Micronav is in the process of establishing a factory acceptance test range adjacent to their facility at Sydney, Nova Scotia. To date, the Antenna Deck and Field Monitor portion of the range has been fully implemented. A temporary far field antenna using a jack-up tower is currently in use while the specification for the far field antenna tower is being finalized. The +40 and -40 antennas are in place but measurements must be made by taking the MLS receiver to the base of the tower. Completion of the range is anticipated in the first quarter of 1990.

Future enhancements to the range that are under consideration include enclosing the antenna deck to protect personnel working with the equipment in inclement weather, implementing an "arc" road to provide a series of points for repeatable measurements at selected azimuth angles, and any landscaping improvements required to eliminate irregularities that may be identified during testing.

This range will provide an efficient, cost effective means to test MLS equipment in simulated installation to prove signal in space operation before the equipment leaves the factory floor.

References

- [1] Skolnik, M.I., Introduction to Radar Systems. New York: McGraw-Hill Book Company, 1962, ch. 7, p.264.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

DESIGN AND ANALYSIS TECHNIQUES FOR MLS SAFETY CRITICAL SOFTWARE

Susan H. Houston, Maryann Vickers
Lisan Corporation
525 School Street, S.W.
Suite 201
Washington, D.C. 20024

Abstract

The Microwave Landing System (MLS) is classified as a safety critical system and must meet exacting system integrity and continuity of service requirements. Integrity, a probability measure, is required to ensure an aircraft on approach will have a low probability of receiving false guidance data. Continuity of service, a reliability measure, is required to ensure an aircraft in the final stages will not be deprived of its guidance information. Embedded software contributing to safety critical operations must exhibit reliability and integrity characteristics such that total system requirements are met. As software reliability cannot be quantified, it is essential that the software be designed and analysed in a manner to meet technical requirements and minimize risk of degrading system integrity due to changes in its environment.

This paper discusses recommended software design and development practices and includes an overview of safety analysis techniques using products of this design process to improve the software design and achieve system integrity goals. Safety analysis is emphasized because its application is relatively new and less familiar than software design practices.

Introduction

The Microwave Landing System (MLS) must meet stringent system integrity and continuity of service requirements. System reliability is currently predicted using the failure rates of the hardware components combined with techniques such as Failure Mode Effects and Criticality Analysis¹. An underlying assumption in such predictions is that all contributions to the quantified system reliability are from hardware failure rates, thus the software contains no errors. It is possible to have random (statistically predicted) hardware failures, but software does not suffer from random failures; every software error is created in the design (or ensuing development) of a system and, therefore, is permanent.

Because MLS performs safety critical functions which may be allocated to software and it is not possible to quantify the contribution of the software to system integrity and reliability calculations, techniques must be developed to provide confidence that the software will not degrade system integrity and reliability. Thus, software must be designed in a manner that ensures it contains no errors. The All-Weather Operations Panel (AWOP) of the International Civil Aviation Organization (ICAO) determined that for MLS, this goal can be reached with an acceptable degree of certainty using and adhering to rigorous software design, documentation, quality engineering, and safety analysis practices.

Good design and development techniques ensure the software meets its specified requirements. The safety analysis ensures the specified requirements are correct and do not contain the ability to degrade system integrity. It requires the application of software safety engineering techniques which are designed to ensure the software contains defensive measures to enhance the safety of the system and that errors which could reduce the safety of the system have been eliminated.

Software Validation

Test is the most common technique to assure the quality of software. Testing is performed to ensure that the software meets the conditions of the requirements specification; that is, the system "does what it's supposed to do". But in the case of software used in safety critical applications, it is implied (if not stated explicitly) that the software should handle all conditions with which it may be confronted and should not permit or cause any dangerous conditions. The system and software requirements specifications may not contain every conceivable fault and the specific requirements to avoid each of them. The purpose of software safety analysis is to identify and minimize the effects of hazardous software faults, when it is not known in advance what their nature is or even whether they exist². Safety analysis seeks to ensure the system "doesn't do what it's not supposed to do".

Software safety analysis incorporates relatively new techniques such as Software Hazard Analysis, Software Sneak Analysis, Petri Net Analysis and Fault Tree Analysis. These analyses, performed in a top down manner from the start of design, greatly reduce the risk of safety critical software errors in the finished product and minimize the cost of changes by discovering problems early in the design process. Since each technique approaches the safety analysis from a different perspective, each one has the potential of uncovering faults not revealed by other methods. No single technique can guarantee correct, error free software. However, a carefully chosen set of techniques can help ensure the development of a software design which will achieve the system integrity and reliability requirements for Category III MLS.

MLS Safety Critical Functions

The MLS is part of an aircraft landing function comprised of a pilot, an aircraft, and the ground based MLS. For an MLS operating in Category III visibility conditions, a failure of the MLS affecting integrity can result in fatality³; the pilot and aircraft will not have information to allow evasive action. Especially for Category III MLS operations, it is essential to achieve strict integrity and continuity of service requirements.

In MLS, the correctness of the guidance radiated to the aircraft is constantly monitored by the executive monitor. The executive monitor detects such conditions as:

- out of tolerance shift of the guidance signal,
- error in data transmission,
- low transmitter power,
- signal format timing or synchronization error

If any such parameters are out of tolerance, the executive monitor automatically shuts down the station and initiates an alarm. If the executive monitor is not functioning properly, the system is shut down. The monitoring and control functions are considered safety critical. If they fail, a hazardous condition may occur. If any safety critical requirements are allocated to and implemented in software, then the MLS contains safety critical software.

ICAO MLS Software Guidance

In 1986, a sub-group of the All Weather Operations Panel of ICAO was formed to determine how to include the contribution of software in system reliability and integrity calculations. This group concluded that, using current state-of-the-art software engineering techniques, the correctness, reliability, and integrity of software cannot be quantified. Without a way to measure the impact of the software on the continuity of service and integrity calculations, the software must be developed in a manner that will provide confidence that it will not degrade system integrity and reliability. The MLS software material recommended by ICAO describes a process to achieve the integrity and reliability requirements by the application of rigorous software design, documentation, quality engineering and analysis techniques⁴. Key features of this process include:

Hardware/Software Partitioning

The partitioning of functions between software and hardware should be clearly specified and justified and should take into account safety criticality, testability, reliability, verification and validation, maintainability, and life-cycle cost.

Software Requirements

Software requirements are based on system requirements. The system and software requirements should be determined carefully and written clearly and unambiguously. Mistakes and omissions in system and software requirements documents are a major source of design errors.

Design and Development Methodology

Development of the software should be a systematic top-down process starting with the formulation of a precise specification based on written system requirements. Each software module should be designed to perform a clearly defined function and have precisely described interfaces. At each stage of the process, the design of each software module should be checked for internal consistency and completeness, and consistency with respect to the designs of higher-level software modules of which it forms a part.

High Order Language Implementation of software should consist of a direct translation of the final, detailed design into a single high order language.

Coding Standards The programming should be performed in such a way that it is possible to trace the correspondence between the code and its design and to verify its correctness with respect to its specification. The coding should conform to strict standards exploiting, to the maximum, the protective features of the high order language.

Code Commenting The code should be interspersed with adequate, clear, and concise comments. The interface of every software module with its enclosing environment should be explicitly stated in a preamble within its code, as well as in the design documentation. Consistency of the code and its comments with the specification and design documents should be checked as each software module is developed.

Software Testing

Software testing must be performed in accordance with a carefully documented test plan. The test plan should include standards and procedures for conducting

all tests, recording the results, and taking corrective action. Software test cases should be defined in parallel with the design process to ensure software testability and adequate test coverage. Software testing should proceed in parallel with the software development in order to reveal faults as early as possible.

Software Documentation

Documentation of the software should be produced in accordance with strict standards as an integral part of the development process. Software documents should have the qualities of traceability, readability, completeness and consistency.

Evaluation and Control of Software Development

To ensure an orderly development of safe, maintainable software that properly performs all functions required of it, activities to evaluate and control the software and its development should be performed concurrently.

Software Quality Assurance (SQA) SQA is the process of assuring adherence to product, development, and maintenance standards.

Configuration Management (CM) CM is the discipline of recording the configuration of a system at discrete points in time in order to systematically control changes to the configuration and maintain its integrity and traceability throughout its life cycle.

Independent Verification and Validation (IV&V) During development, IV&V is performed at every phase to ensure that the product is properly constructed and consistent with its specification. The distinction between SQA and IV&V is that SQA assures adherence of the development methods and product standards to prescribed standards, while IV&V is performed to ensure that the product meets its functional requirements.

System Safety Analysis

System safety analysis should be performed to uncover faults, including omissions from requirements and behavior in degraded modes, that compromise system safety. Software safety analysis techniques are used to assure that a system does not cause hazards while performing its mission. By providing a different perspective from other test and analysis techniques, it may also uncover functional inadequacies. The safety analysis should be performed during both design and development.

Using the software development methodology recommended by ICAO should result in software that is reliable and maintainable. In order to have confidence that the software does not degrade system integrity, the safety analysis recommended by ICAO must be performed. The following sections explain the roles of safety analysis in the design context.

Safety Analysis Performance

Comparison of FMECA and Safety Analysis

Failure Mode, Effects, and Criticality Analysis The FMECA is intended to verify design fault tolerance by identifying each failure mode and determining the system's reaction. It ranks the effect of each failure in terms of its severity. For hardware failures, a criticality number is calculated based on the failure probability combined with its relative severity. The criticality analysis requires a failure rate which is not available at the functional (or

software) level. Results of FMECA are used as rationale for design changes, the addition of fault tolerance and recovery mechanisms, functional partitioning, and testability issues.

Safety Analysis The safety analysis differs from FMECA by concentrating on safety related problems exclusively. The safety analysis is an analysis format and typically combines many analysis techniques in its application. It identifies those hazards that can affect safe system performance due to either a failure (degraded operation), design flaw, or specified requirement that allows the system to operate in an unsafe mode. The safety analysis considers each hazard, its possible causes, the most critical mode of operation during occurrence, methods of avoiding or minimizing danger, and a severity ranking of the effects.

Combined FMECA and Safety Analysis Techniques The safety program should be designed to complement FMECA activities. As results are derived from system and piece-part (hardware component) FMECA's, all failure modes that impact software should be input to the respective software safety analysis activities. Using a suite of safety analysis techniques that together analyze all appropriate aspects of the software design will fulfill safety analysis requirements as well as verify FMECA results. The U.S. Military Standard on System Safety Program Requirements⁵ gives guidance for a complete system safety analysis program including software safety analysis.

The U.S. Federal Aviation Administration (FAA) specifications require a system level FMECA on all functions, and a piece-part (chip level) FMECA to be performed on all modules contributing to integrity and continuity of service⁶. The results are used in combination with analysis to show that integrity and continuity of service requirements are met. Ordinarily, a FMECA would not be sufficient to analyze system integrity requirements. However, the FAA requires FMECA failure modes to be considered in the absence of correct monitor operation. The consideration of multiple failure modes mimicks the safety analysis techniques used to ensure design behavior under hazardous MLS operating conditions (ie, loss of monitor). The system safety analysis can be tailored to complement the FMECA by focusing its review on the translation of requirements into the design concept, identifying interdependencies among functions to assure no unsafe conditions are present, and identifying the system level hazard conditions, such as unsafe mode transitions, human interface hazards, and monitor failure modes.

MLS System and Software Safety Program Activities

The software design and development process is comprised of the following activities⁷:

- System Requirements Analysis,
- Software Requirements Analysis,
- Preliminary and Detailed Design,
- Coding and Software Unit Testing,
- Software Unit/Module Integration and Testing,
- Software Configuration Item Testing,
- System Integration and Testing.

Structured software design activities produce products which will be used for safety analysis. Typically, these include graphical representations of the design such as data flow diagrams, functional flow diagrams, control flow diagrams and state transition diagrams. Computer aided software design (CASE) tools are available to develop these products and perform a variety of analyses. Designers use these products to

make sure the design performs as intended; safety analysis uses them to ensure the design does not perform unintended and hazardous functions. The following sections discuss the system and software related tasks to be incorporated in an MLS system safety program plan.

1) Identify System Safety Requirements and System Level Hazards System safety requirements should be explicitly stated in the system specification. In the case of MLS, integrity and continuity of service requirements are specified. The MLS technical requirements contributing to integrity and continuity of service must be identified. From these, all hazards that can cause unsafe operation are identified including monitor actions, environmental conditions, maintenance operations, operator error, and specification inconsistencies and unsafe modes. For MLS, functions to identify should include those which may cause⁴:

- Significant increase in path following error,
- Error in basic datavord 2, minimum glide path,
- A synchronization error resulting in overlap,
- Unacceptable increase in control motion noise.

2) Perform System Safety Analysis As part of system requirements analysis, a safety analysis on the hazards identified in 1) is performed to identify critical functions for further analysis and design considerations, and specification changes if necessary.

The system safety analysis should use the functional flow diagrams (or equivalent) produced during system requirements analysis to determine if functional paths exist to the hazardous states and inputs identified. The fault tree analysis and petri net analysis are appropriate at this level.

3) Perform System Level FMECA The system level FMECA is performed on the entire system. All possible effects for each failure mode is noted, along with any detection and correction methods. Frequently, more than one effect per failure mode exists depending on the mode of operation of the function under consideration. All such effects should be identified to assure consideration of the most critical ones. At the system level, failure modes usually result in loss or degradation of performance parameters. The benefits to be gained at the system level are the identification of potential single point failure modes, and the ability to view the system from the perspective of degraded modes to facilitate development of fault tolerance, detection, and recovery schemes. It also provides the basis for functional partitioning and subsequent analysis.

4) Functional Partitioning Using the results of system analyses, the design team will partition functions among hardware and software. Partitioning should take into consideration safety critical factors, testability, fault detection schemes, fault recovery schemes, modular independence, and ease of implementation. The functional flow diagrams should be updated at this point to reflect partitioning.

5) Analyze Software Safety Critical Requirements As part of software requirements analysis, a careful analysis of the safety critical requirements allocated to software should be performed. The feasibility of the implementation should be assessed and the testability of each requirement determined. The qualification method for each requirement should be identified at this point. Software safety requirements should be explicitly documented in the software requirements documentation. Safety analysis need only be performed on safety critical software.

6) Perform Software Safety Analysis During Design The top level design is analyzed to determine if all safety requirements allocated to software have been incorporated into the design. The safety requirements should be allocated to independent modules to reduce risk and to facilitate verification and test. The system level FMECA identified failure modes and effects that occur in functions now allocated to software. These entries will be updated and included in a software functional safety analysis to verify the software's reaction as identified in the system FMECA.

As the detailed design is developed, software safety analysis is performed using the updated data and command flow diagrams to analyze the design's ability to react to any unsafe states the system could enter. Software hazards include undesired state changes, incorrect timing of events, failure of an event to occur, and incorrect input to an event. Based on the results of the analyses, the design is modified to eliminate as many unsafe states as possible, keeping only those possible unsafe states necessary to provide adequate functional and performance behavior. This analysis is updated to include the code as input.

7) Safety Analysis During Implementation At the code level, much more detailed failure modes can be considered. Failure modes that can cause suspension of normal program execution include :

- Wrong initialization or reset of variables,
- Incorrect invocation of subroutines,
- Variables out of range,
- Incorrect timing,
- Memory inavailability,
- Incorrect instruction,
- Mathematical mistakes such as division by zero.

Exception handling code to identify these conditions should be added where possible. Computer Aided Software Engineering (CASE) tools are available with capability to assist in static and dynamic analyses including tracking of variables, input/output and acceptable ranges, subroutine invocation, timing, unused code, test case coverage, and other software attributes.

The Software Sneak Analysis can be used during code level safety analysis to identify all possible paths of safety critical software operation. This is important due to the possibility of incorrectly invoking software that may allow hazardous operation. The software sneak analysis uses safety critical code or Program Design Language (PDL) level data and control flow diagrams of the software paths. From these, all path traversals are determined and reviewed for unwanted state transitions. This analysis will also verify safety critical code independence.

8) Software and System Testing The system should be tested under various stress and degraded modes to ensure that the protective features have been implemented correctly as a result of analysis. The Radio Technical Commission for Aeronautics DO-178A gives detailed guidance on testing goals relative to software and system testing⁸. In summary, testing should span beyond requirements validation to include the following:

- The unit level tests should verify the code performs only its intended functions and contains logic to terminate loops, handle all cases of input, and handle missing and out of range input;

- Module integration tests module interface characteristics such as data and control flow, subroutine argument types and ranges, illegal outputs, timing, sequence of events, and module independence;

- System integration looks for incorrect interrupt handling, timing constraints and timing errors, software reactions to hardware failures and transients, built-in-test failure detection, start-up and transient load handling, bus resource contention, and hardware/software interface validation.

Test cases should be selected to use inputs from FMECA and safety analysis; testing should include simulation of hardware faults and real-time execution.

Summary and Conclusions

Safety critical software requires special considerations during design and development. These include rigorous requirements analysis and documentation, top down design, thorough software documentation, adherence to SQA and CM practices, and performance of safety analysis with FMECA to identify and account for all the ways the software can perform in its environment and react to hardware transients and failures. Testing must be designed to exercise the defensive mechanisms built into the software as a result of analysis. Many techniques exist for performing thorough software safety analysis; contractors should choose a combination of techniques to complement each other and be an integral part of their design and development processes. Automated tools should be used to assist in the design process and facilitate software analysis and test. Applying these methodologies to safety critical software is the optimal way to achieve high system integrity and continuity of service and ensure proper MLS software behavior in the system environment.

References

1. Military Standard 1629A, Notice 1, "Procedures for Performing a Failure Mode Effects and Criticality Analysis," U.S. Department of Defense, Washington, D.C., 7 June 1983.
2. A. L. McFarland, Y. S. Hoh, and G. A. Dorfman, Procedures for Cause Oriented Analysis and Consequence Oriented Analysis. U.S. Department of Transportation Report, DOT/FAA/PM-86/21, June, 1986, pp. A-2 to A-3.
3. T. H. Bohr, et al., Report presented at the Twelfth Meeting of the All-Weather Operations Panel, International Civil Aviation Organization, Montreal, Canada, 9-27 November, 1987. pp. 2H-1 to 2H-5.
4. *ibid*, pp. 2I-1 to 2I-6.
5. Military Standard 882B, Change Notice 1, "System Safety Program Requirements," U.S. Department of Defense, Washington, D.C., 9 July 1987.
6. Federal Aviation Administration Specification FAA-E-2721/11, "Microwave Landing System Ground Equipment General Requirements," U.S. Department of Transportation, 19 June 1986, pp. 33-34.
7. Military Standard DOD-STD-2167A, "Defense System Software Development," U.S. Department of Defense, Washington, D.C., February 1988, p. 9.
8. Radio Technical Commission for Aeronautics RTCA/DO-178A, "Software Considerations in Airborne Systems and Equipment Certification," Washington, D.C., March, 1985, pp. 24-28.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

DME GROWTH ELEMENTS AND THEIR USE WITH MLS

A. Becker

German Aerospace Research Establishment, (DLR), Institute for Flight Guidance,
Braunschweig, Federal Republic of Germany

1. Introduction:

In 1978 the TIME Reference Scanning Beam System was selected by the All Weather Operations Divisional Meeting (AWOD) of the ICAO as the new MLS. It provides vertical as well as horizontal guidance information in an approach sector of ± 60 degrees. The interception of this sector, however, must be performed by the use of other navigational aids, e. g. VOR. In order to improve the capability of MLS a recommendation was given in the AWOD report [4] towards the development of a 360° degrees azimuth element to be integrated with the MLS. A corresponding time slot was provided in the TRSB signal format. But, as a cost effective alternative, investigations on a DME-based element were encouraged using the techniques of the DLS, the FRG-proposal for the MLS. In this context the concept of the DAS was developed by the SEL company. A different DME-based concept was investigated by the DLR. In this paper, this concept will be explained in some detail, after a short discussion of the growth potential of the DME. It will be shown further, that this concept additionally can be upgraded to a roll guidance and airport surveillance system.

2. The Growth Potential of the DME

The DME is a well-proved component of air navigation systems that was introduced in 1959 and standardized internationally since then. The standards now are extended to the DME/P, a component of the new MLS [5]. Both, the DME and DME/P, will be used far beyond the year 2000.

It is well known that in most applications the pulse space of the DME is used to only a small degree. Therefore, additional functions can be integrated into the system by inserting further pulses into the signal format. These pulses can be modulated with the desired information (e.g., by pulse position modulation). Moreover, additional information can be derived from the conventional DME signal in space. These are the azimuth and the elevation angles, respectively, of the received DME interrogation wave on the ground that indicate the direction to the aircraft. All of these additional functions can be obtained without loss of compatibility with the standard DME. The requirements for the pulse shape and rf spectrum [5] are maintained. Some of these additional functions were already realized with the DLS and DAS. A set of possible growth elements is described in Refs. [3, 6]. In the following, after a short review of the DME and the scanning DME, those growth elements are shortly described which are candidates for a 360° AZ-subsystem of the MLS. Finally, a preferred solution is selected and its extension to a roll guidance and airport surveillance system is indicated.

2.1 Distance Measuring Equipment (DME)

With DME the sum of the travel times of an interrogation pulse pair from the aircraft to the ground transponder and of an reply pulse pair from the transponder to the aircraft is measured. From this measurement the slant range of the aircraft to the ground transponder can be derived onboard the aircraft. Thus, an air-derived line of position is available onboard.

2.2 Scanning DME

With the scanning DME a number of ground transponders are interrogated consecutively by switching the interrogation frequency to the downlink channels of the allocated ground transponders. The different slant range measurements are tracked onboard independently (up to five trackers). Thus, an over-determined position can be get with high accuracy in the ground plane. The altitude calculation, however, must be supported by a barometric altitude measurement. The scanning DME measurement again is air-derived, for the same reasons as with DME.

2.3 Three-Way DME (TW-DME)

The TW-DME is an extension of the standard DME. It provides the slant range information on the aircraft as well as on the ground. The principle is shown by Fig. 1. The figure also shows the integrated data links, uplink as well as downlink [1,3,6].

In Fig. 1, the airborne DME interrogates the ground transponder (route 1). The ground transponder transmits a reply pulse pair back to the aircraft (route 2). Thus, the pulse travel time (which is proportional to the slant range) can be measured onboard the aircraft. This is the basic function of the standard DME. In the case of the TW-DME, however, the reply pulse pair is supplemented by a third pulse, the so-called marking pulse. Simultaneously, a time counter is started on the ground. When the onboard equipment detects the marking pulse, the airborne transmitter sends a further pulse pair (route 3). After detection of this pulse pair at the ground station, the counter is stopped. The slant range information is thus also available on the ground. In order to distinguish between the pulse pairs of routes 1 and 3, the pulse spacing is different. A reply of the transponder to a received pulse pair will be initiated only by the ICAO standardized pulse spacing on route 1. Thus, the compatibility with the standard DME is maintained.

The TW-DME with its extended capability compared to the standard DME generates an air-derived line of position onboard the aircraft as well as a ground derived line of position on the ground. The integrated data links shown in Fig. 1, uplink as well as downlink, use the principle of pulse position modulation for encoding the data to be transmitted. In the uplink the DME reply pulse pair is used as the time reference. The distance Δt_{az} encodes the ground derived azimuth angle. Δt_{az} ranges from $\Delta t_{az,0}$ to $\Delta t_{az,max}$. In the downlink the same principle is used. Here, the airborne reply on route 3 serves as the time reference. The data "identification" and "altitude" control the positions Δt_{id} and Δt_{alt} of two consecutive pulse pairs. In the uplink as well as in the downlink further data windows for additional data can be provided if required.

2.4 Trilateration Three-Way DME (TTW-DME)

The TTW-DME is an extension of the TW-DME. It has the capability of multiple slant range measurements on the ground and thus the capability of a complete position determination of the aircraft by trilateration using only one DME channel [6, 7].

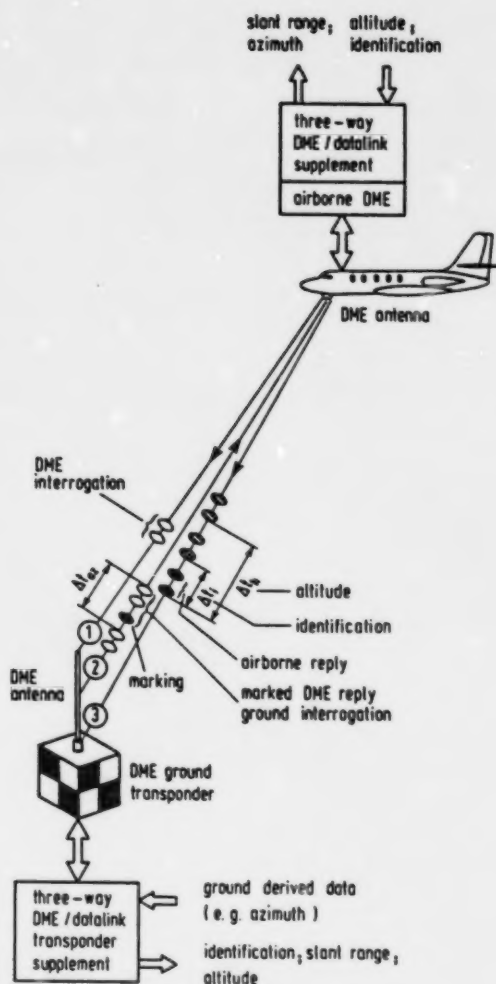


Fig. 1 Principle of the three-way DME with integrated air-ground and ground-air selective data links

The principle can be taken from Fig. 2. The aircraft and the central ground station are linked via a TW-DME. The three routes are indicated by 1, 2 and 3. In addition, two substations are positioned at exactly known locations having the distances a_1 and a_2 from the central station. These substations consist mainly of DME receivers which are tuned to route 3 of the TW-DME. Thus, the pulse pairs of route 3 not only generate trigger pulses in the central ground station, but also in both substations. These trigger pulses are transmitted by a wide band rf-link to the central ground station. The pulse travel time measurements on the ground now can be made via three different paths.

- 1) Route 2 + route 3, (TW-DME) $\rightarrow d_1$.
- 2) Route 2 + route 3, + a_1 $\rightarrow d_2$.
- 3) Route 2 + route 3, + a_2 $\rightarrow d_3$.

Measurement 1) provides the slant range d_1 . Measurements 2) and 3) can be used for the determination of d_2 and d_3 , respectively, by taking into account the known distances a_1 and a_2 , as well as the already measured slant range d_1 . From these slant ranges, the position of the aircraft is calculated with respect to

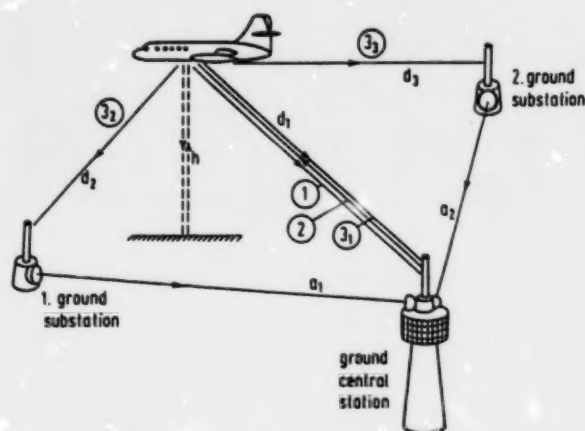


Fig. 2 Principle of the trilateration three-way DME (TTW-DME)

the central ground station. The z-coordinate of the aircraft at low elevation angles can be obtained with only poor accuracy. Therefore, a direct measurement of the altitude (barometric or radio) is provided on the aircraft. This altitude is transmitted to the ground by the integrated air-to-ground selective data link (see Fig. 1)

The position determination of the TTW-DME is ground derived. For navigational usage of this ground derived information a ground to air selective data link must be available. On board, the complete position can be calculated from the transmitted azimuth angle in combination with the air derived slant range measurement and the barometric altitude. Obviously, the azimuth angle can also be used with MLS for the goal envisaged in this paper.

2.5 Differential DME (Δ -DME)

The Δ -DME [2, 3] shows a close relation to the TTW-DME. This follows from a comparison of Fig. 3 to Fig. 2. The substations in both systems are nearly identical. The main difference, however, is the principle of position determination. In contrary to DME the Δ -DME measures the difference of the downlink pulse travel times between two ground stations, i. g. the central station and one substation in Fig. 3. The area of position of the aircraft is a hyperboloid with both stations as focus points. A second pulse travel time difference, i. g. measured by the central station and the other substation, generates a second hyperboloid. The intersection of both areas is a hyperbola. If the barometric altitude of the aircraft is transmitted to the ground via a data link (see subchapter 3) the intersection of the hyperbola with the altitude plane can be calculated on the ground. Thus a ground-derived position determination of the aircraft is obtained. When a third ground substation is installed, a redundant position determination is possible with a corresponding increase in accuracy and integrity.

With respect to the different principle of measurement the error behavior of the Δ -DME differs from that of the DME. This statement is true for the basic measurement as well as for the geometric dilution of precision (GDOP). The error comparison for both systems concerning the basic measurement is illustrated by Fig. 4 and Table. 1. The line of position in the two-dimensional case is a circle with DME and a hyperbola with the Δ -DME. The area which is limited by the 95%

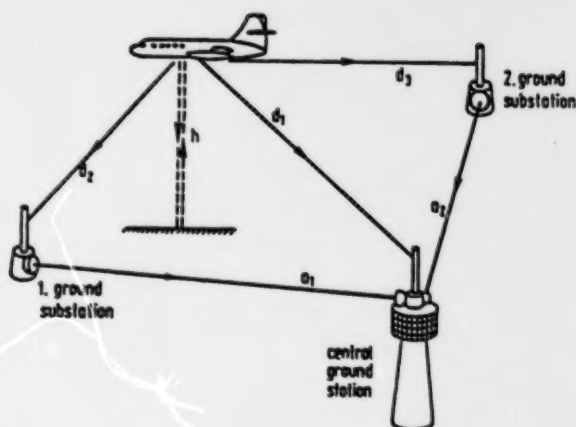


Fig. 3 Principle of the Differential-DME (Δ -DME)

probability curves is illustrated in Fig. 4. The best accuracy with the Δ -DME can be obtained on the baseline between the two groundstations. Table 1 indicates on the left side the order of the error contribution of different effects corresponding to the Annex 10 document of ICAO ([5]). The RMS addition gives a 2σ DME value of 23,9 m which is in good correspondence to

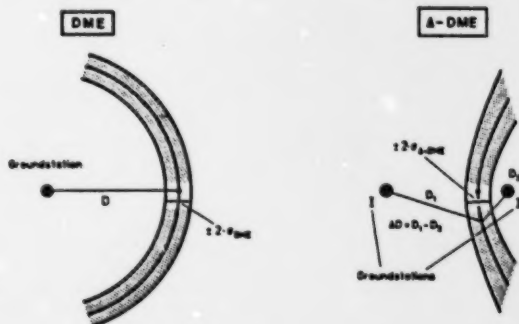


Fig. 4: Comparison of Standard Deviation in the Case of DME and Δ -DME (two dimensional)

the performance measured at DLR. The right side of Table 1 shows the Δ -DME error components. With respect to the differential measurement principle the airborne transmitter does not contribute to the error. The time of transmission of the pulse is of no importance. On the ground two receivers are used in the case of Δ -DME. With the assumption that the DME transponder error of 10 m consists of statistically independent equal contributions of the receiver and the transmitter, an error of 7 m is assigned to receiver 1 and receiver 2. With DME the downlink wave takes the same path as the uplink wave. Therefore, the same multipath error must be assigned to uplink and downlink. With Δ -DME, however, only the downlink exists using two different groundstations with a large geometrical separation. It is unlikely that both paths are contaminated by multipath effects at the same moment. Therefore, only one specular multipath error is taken in the table. The same argumentation can be applied to the garble effect. The RMS addition of the error component results in a 2σ Δ -DME error of 15,6 m which is considerably lower than the corresponding DME-error.

Error Components

DME/P (see Annex 10)		Δ -DME	
Transponder	10 m	Receiver 1	7 m
Interrogator	15 m	Receiver 2	7 m
spec. Multipath (downlink)	10 m	spec. Multipath	10 m
spec. Multipath (uplink)	10 m	diffuse Multipath	3 m
diffuse Multipath	3 m	Garble	6 m
Garble	6 m		
$2\sigma_{DME} = 23,9 \text{ m}$		$2\sigma_{\Delta-DME} = 15,6 \text{ m}$	

Table 1: Error Budget of the Δ -DME in accordance to the ICAO DME Error Budget (Annex 10).

For a complete position determination in three dimensions the intersection of three areas of position must be calculated. In the plane only two lines of position must be intersected. The accuracy of the position determination depends on the accuracy of the basic measurement (see Table 1) and the intersection angle. Rectangular intersections minimize the error while small intersection angles cause high errors. In practice, small angles often can not be avoided in certain areas. This effect is described by the GDOP (geometric dilution of precision). In Fig. 5 this effect is illustrated for the Δ -DME with three regularly positioned groundstations. The position determination obviously is accurate within the triangle formed by the groundstations. Outside this triangle, however, the radial error increases progressively with distance while the lateral

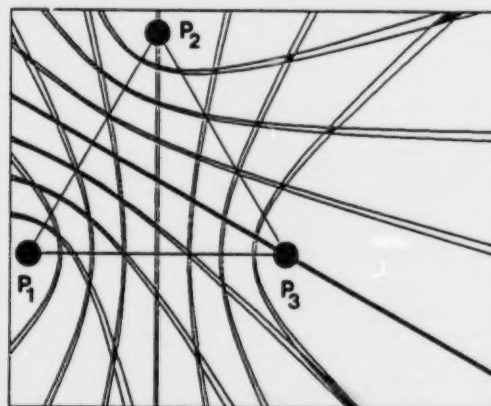


Fig. 5: Illustration of GDOP for Δ -DME in the ground plane

error is acceptable. It can be shown that the system at large distances behave like a pure angle measurement system. It can further be shown that the accuracy of the angle measurement is superior to the corresponding angle measurement accuracy of three DME stations using the same baselength. This is true due to the fact that the basic accuracy of the Δ -DME is higher than the basic accuracy of the DME (see Table 1). In applications which require a good angular accuracy only, the Δ -DME, therefore, is superior to the TTW-DME. This is the case with a 360° AZ-subsystem for MLS. Thus, for this application, the selection of the Δ -DME is justified.

3. A 360°-Azimuth Guidance Subsystem for MLS

According to the recommendations of the AWOD-report of the ICAO [4] a 360° azimuth guidance subsystem as supplement of MLS can be based on the TRSB or the DME. A number of advantages support a selection of the Δ -DME into which downlink and uplink data links are integrated.

The installation is illustrated by the scenario in Fig. 6 showing the layout of the airport Munich II. The airport is under construction at this time. The central ground station is assumed to be placed at the tower, while four substations are placed at the corners of the airfield. Thus, the position measurement is over-determined to a high extent allowing up to two substations to be shadowed or contaminated by strong multipath. While the central ground station (c) contains a complete DME transponder, the substations (S₁) to (S₄) consist of DME receivers only with r.f. wideband connections to the central station.

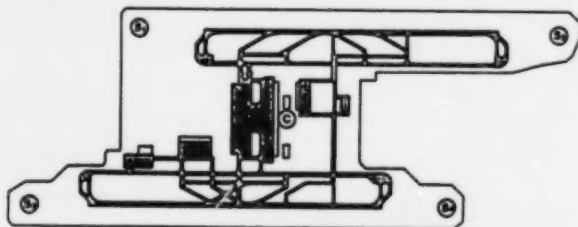


Fig. 6: 360°-Az-guidance subsystem based on Δ -DME in the Munich II scenario

On board the approaching aircraft a DME/P or a DME is needed supplemented by a data coder (downlink) and a data decoder (uplink). Corresponding coding and decoding circuits must be implemented in the central ground station.

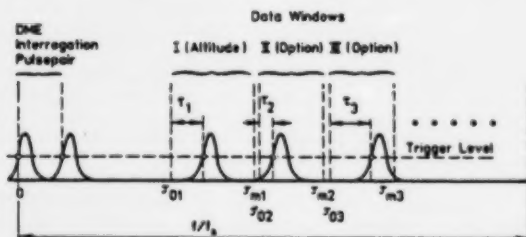


Fig. 7: Downlink Selective Data Transmission

The selective downlink data transmission is illustrated by Fig. 7. It is done by pulse position coding of the information to be transmitted, i.e. the barometric altitude. The DME interrogation pulse pair is used as the reference for the data pulses. The time window for the altitude data pulse ranges from τ_0 to τ_m . τ_1 represents the actual altitude measurement. Further data windows can be provided if required.

The selective uplink data transmission is illustrated by Fig. 8. The coding principle is the same as with the downlink. However, in the uplink pulse pairs like the DME reply pulse pair are used for en-

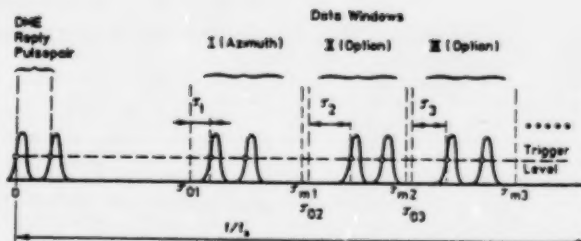


Fig. 8: Uplink Selective Data Transmission

coding of the data information in order to improve the rejection capability against wrong pulses from other sources. In the downlink this measure is not needed because the data transmission is N-1-fold redundant when N groundstations are used. The azimuth information evaluated in the ground computer is transmitted to the corresponding aircraft in data window I. Further windows can be provided if required.

4. Prospective View

The system described in chapter 3, provides the aircraft with a 360°-azimuth information. When hardware components of DME/P technology are used, the 2 σ -error can be kept below 0,2° for the scenario of Fig. 6 [8]. This accuracy is quite satisfying for the envisaged application. When the aircraft is landed, however, a further application of the system is evident. From Fig. 5 follows that the complete position of the aircraft can be detected with high accuracy. By using the full data link capacity the information needed for a roll guidance and airport surveillance system can be made available on the ground as well as in the aircraft [8].

5. References

- [1] Ullrich, R. (1979). Vorschlag für ein Dreiweg-DME-Verfahren in Verbindung mit dem Mikrowellen Landesystem DLS, DFVLR Institut für Flugführung, Institutsbericht IB 153-79/05.
- [2] Schnittger, H. J. (1980). Genauigkeitsuntersuchungen zum Differenz-DME. DFVLR, Institut für Flugführung, Institutsbericht IB 153-80/20.
- [3] Becker, A. (1980). Conceptual Considerations on DME Growth Potential. DFVLR Institut für Flugführung, Institutsbericht IB 153-80/22.
- [4] ICAO, Report of the All Weather Operations Divisional Meeting, (AWOD) Montreal, April 1978
- [5] ICAO, (1985). Annex 10 to the Convention on International Civil Aviation, vol. 1.
- [6] Becker, A. (1986). Integrated navigation communication, and surveillance systems based on standard distance measuring equipment. *Journal of Aircraft*, 23, No. 5
- [7] Schnittger, H. J. (1984). Genauigkeitsuntersuchungen zu einem auf dem Dreiweg-DME basierenden Trilaterationsverfahren. DFVLR, Institut für Flugführung, Institutsbericht IB 112-84/29.
- [8] Becker, A. (in preparation). A Rollguidance and Airport Surveillance System, based on Differential DME. DLR, Institute for Flight Guidance, Braunschweig, DLR Report.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

DME/P TRANSPONDER REPLY DELAY ACCURACY - FACTORY VERIFICATION AND CALIBRATION

Elven H. Nisson

E-Systems Inc., Montek Division
2268 South 3270 West M/S 23
Salt Lake City, UT 84119 U.S.A.

Abstract: To avoid complication of site installation and improve system test flow, it is critical to determine the reply delay performance of the Precision Distance Measuring Equipment (DME/P) portion of the Microwave Landing System (MLS) during testing on the factory floor. This paper covers methods of measuring and demonstrating that the critical reply delay performance parameters are met during factory testing, including calibration of the test equipment for the measurements and calibration of the DME/P as part of the testing process. The paper concludes with comparisons of DME/P performance at various stages from initial calibration and test at the factory through site installation and field testing.

Reasons for Precise Measurement Calibration

When a DME/P transponder is built, there may be delay error buildups in the various assemblies of the DME/P transponder that result in slight biases in the reply delay of the system. These biases may add up to no more than 15 to 20 nanoseconds. This is, however, one fourth to one third of the entire delay error allowed for the DME/P transponder. Time drifts due to changes in environmental conditions could move this delay error beyond acceptable limits. Calibration of the DME/P transponder becomes an important issue in the delivery of hardware. The point at which the calibration is performed is also an important issue. Calibration of the basic DME/P transponder reply delay during factory testing provides several advantages over calibration at other points in the site preparation process.

Simplified Installation Procedure

The first advantage is simplification of the site installation procedure. The installation of a DME/P transponder in conjunction with an MLS system involves initializing the DME/P transponder equipment to the site specific conditions where the transponder is being installed. A DME/P transponder calibrated at the factory needs only to be programmed with the specific site parameters based on the cabling used and expected echo and multipath conditions. The precalibrated and newly site programmed DME/P transponder takes into account the unique siting conditions and adjusts the calibrated reply delay accordingly to maintain a properly centered reply delay and monitoring accuracy.

Reduction of MLS System Test to Interface Test

The second advantage is a reduction in the complete MLS system test requirements. The MLS system test prior to final shipment can be held to a validation of the interface cabling and message handling capabilities of the component pieces. This is due to complete performance calibration and testing of the DME/P transponder during lower level tests. The design of the E-Systems DME/P transponder system is such that the reply delay performance of the DME/P transponder is not dependent on the operation of other parts of the MLS system.

Simplified and Repeatable Recalibration After Repair

The third advantage is a simplified and repeatable recalibration after a repair. When a repair has been completed on a DME/P transponder, a procedure similar to the factory calibration can be used to check and adjust, if necessary, the reply delay calibration factors set at the factory. This calibration is accomplished with test equipment designed to provide reply delay data for the DME/P as measured from the output of the transponder cabinet.

Reply Delay Measurement Methodologies

The measurement of the DME/P transponder reply delay in a test situation requires higher measurement accuracies, more stringent pulse detection methods and higher accuracy measurement methods than will be seen in normal operation. These more stringent requirements

allow for precise testing and calibration of the DME/P transponder. The aircraft distance measurement is based on measuring the round trip delay of an interrogation from the aircraft to the DME/P transponder and back to the aircraft. When the aircraft is in the Initial Approach (IA) mode, it interrogates at a rate of 16 Hz. In the Final Approach (FA) mode, the aircraft interrogation rate is 40 Hz. The measurement of reply delay in a test environment employs repetitive signal averaging of secondary measurements to improve the accuracy of each reply delay measurement. IA mode and FA mode test interrogations that are not used for reply delay measurements are interspersed between the IA mode and FA mode interrogations that are used to determine reply delay performance. This is done to complete the precise measurement of the interrogation and reply pulse shapes. These precise pulse shape measurements are used to improve the measurement of reply delay as defined in the international specifications [1], [2]. These measurements are accomplished by using a precision detector and digital to analog (D/A) converter to precisely determine specific voltage levels at which measurements occur as explained in the following paragraphs.

Continuous Data Tracking

The determination of reply delay error is based on a large sample of processed transponder reply delay measurements. The measurement of reply delay in the test environment collects a continuous sample of reply delay data for processing. To maintain the integrity of the measurement processing, the data are collected at the specified interrogation rates of 40 Hz for FA mode and 16 Hz for IA mode [1].

Interrogation and Reply Peak, 50%, 30% and 5% Points: The first of the required measurements to improve the accuracy of the factory measurement of transponder reply delay is to determine the peak detected voltage of the interrogation and reply pulses. Once the interrogation and reply peak voltages are determined, the peak values are maintained on a continuous basis. After the peak voltage has been determined, the 50%, 30% and 5% voltage points can be determined separately for the interrogation and reply pulses. These values are also maintained on a continuous basis throughout the reply delay measurement process. These voltage points are used in the measurement of reply delay.

Reply Delay Measurements from 50% and Virtual Origin Points: The measurement of IA mode reply delay is accomplished in a straightforward manner as described in the international specification documents [1], [2]. The time delay from the 50% point on the rising edge of the first interrogation pulse to the 50% point on the rising edge of the first reply pulse is measured. This measured value is used as the raw measurement of IA reply delay.

The measurement of FA mode reply delay is somewhat more complicated to perform. The measurement is specified to take place from the virtual origin of the first interrogation pulse to the virtual origin of the first reply pulse. Since this cannot be directly measured as the virtual origin is a calculated rather than an actual point, the measurement must be made at an actual point and then the virtual origin calculated from the measurement.

The virtual origin is defined as the 0.0 voltage point on the line extended downward and passing through the 5% and 30% points on the leading edge of the pulse. The same points are used in the measurement of reply delay. The test equipment maintains a moving average of the time delay between the 5% and 30% points on the rising edge of the interrogation and reply pulses. The basic FA mode reply delay measurement is completed by measuring the delay from the 30% point on the rising edge of the first interrogation pulse to the 30% point on the rising edge of the first reply pulse. The basic reply delay measurement is then adjusted by six-fifths of the 5% to 30% rise time value (0% to 30% time is six-fifths of 5% to 30% time) from both the interrogation and reply pulses. This converts the basic reply delay

measurement into a correct virtual origin to virtual origin measurement.

Conversion of Raw Data to Useful Performance Data

After the raw transponder reply delay measurements have been completed, the data must be processed to determine compliance with defined requirements. The processing of data can be completed on a real time basis between interrogations or post-processed after the collection of the raw data. All data shown throughout this paper are post-processed data.

Collection of Continuous Raw Data Sample Set: The amount of data, desired to be used in the performance measurement vary depending on the external conditions imposed on the performance measurement. A typical number of samples that will be available to an aircraft on landing will be 5,760 IA mode interrogations and 7,80 FA mode interrogations. These numbers are based on an average landing airspeed of 150 knots over the entire 22 nautical miles (nmi) straight in landing approach with a transition from IA mode to FA mode beginning at 8 nmi range and completing at 7 nmi range. The data for measurement are collected in raw form as basic measurements and offset factors for later calculations. The data are collected continuously in the same manner as an aircraft on a landing approach. After the raw data are collected, they are processed for evaluation. Figure 1 shows a time plot graph of FA mode raw data collected from a DME/P on a factory floor.

FA Data Processing - Raw Reply Delay Performance Data

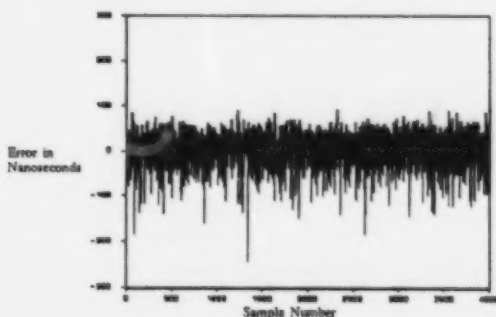


Figure 1

Filtering Raw Data to Selected Filter: The determination of the performance of a DME/P transponder is specified in terms of Path Following Error (PFE) and Control Motion Noise (CMN). These two quantities have been defined to model the control surfaces and actions of an aircraft during a landing operation. Figures 2 and 3 show time plot graphs of FA mode PFE and CMN data from the raw data shown earlier.

FA Data Processing - PFE Filter Output

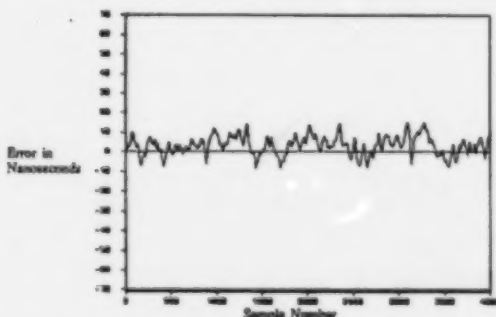


Figure 2

Determination of Error: A precise procedure for the determination of the error and the checking of limits has been defined for the DME/P transponder. A time window of filter output data is

processed to determine the transponder performance. The filter value of a window is the point at which 95% of the magnitudes of the data in the window are closer to 0 error. With the actual error an unsigned magnitude rather than a signed value, the absolute value of the filter output is used to determine the error value. This filter value is then compared to the defined limits for the specific filter. The window is then moved across the filter data one point at a time and new 95% points are determined at each new window point. This is easily accomplished by maintaining a sorted list of the absolute values of the filter output. Figures 4 and 5 show time plot graphs of the absolute values of the FA mode PFE and CMN filter data. This sorted list is then updated with each new addition from the filter output and a new 95% point is determined. Figures 6 and 7 show time plot graphs of the final 95% values for FA mode PFE and CMN filter data.

FA Data Processing - CMN Filter Output

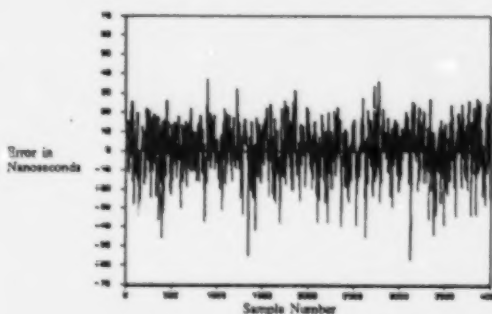


Figure 3

FA Data Processing - Absolute Values of PFE Data

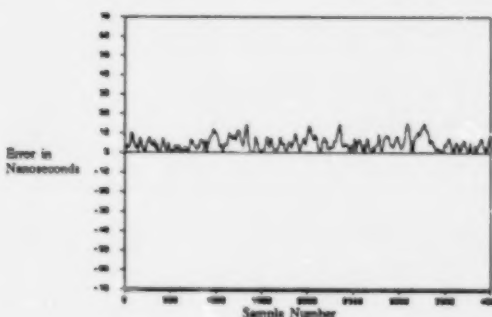


Figure 4

FA Data Processing - Absolute Values of CMN Data

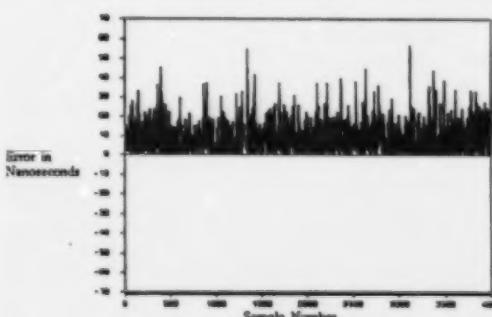


Figure 5

FA Data Processing - 95% PFE Performance

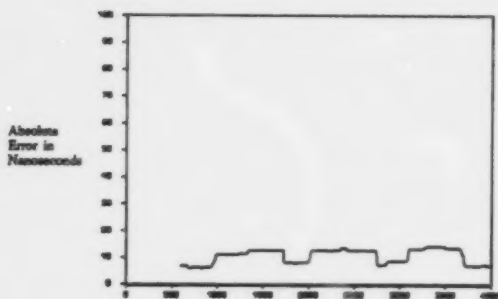


Figure 6

FA Data Processing - 95% CMN Performance

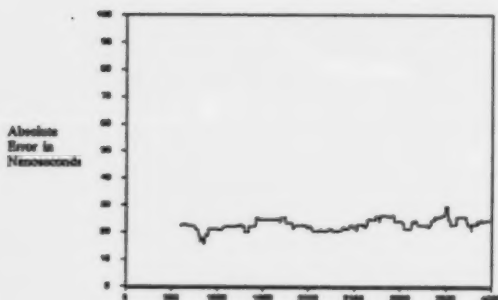


Figure 7

Pitfalls and Errors Possible in Measurements

A system is used to measure time delay between two signals can have a number of error sources. Of all the possible measurement errors, two measurement error sources are discussed that pertain specifically to the measurement of reply delay in a DME/P system.

Reply Delay Measurements Based on Interrogation and Reply Pulse Triggers: Perhaps one of the easiest mistakes to make is to use the triggers for the interrogation and the reply pulse pairs for the reply delay measurement. The logic is based on the expectation that the variation in delay from trigger to generation of the Radio Frequency (RF) signal will be small in comparison to the variation in delay due to the detection of the interrogation. This assumption allows for measurements of reply delay to be made between the interrogation and reply triggers with appropriate offset calibration factors available to adjust the measurement to the desired measurement point.

This measurement method works well as long as the pulse shape of the interrogation and reply are maintained within tight standards. Figure 8 shows a COS/COS³ (DME/P) pulse superimposed on a Gaussian (Distance Measuring Equipment / Normal (DME/N)) pulse. The pulses shown are time equivalent at the 0.5% pulse amplitude point. This is a simulation of a COS/COS³ and a Gaussian pulse being triggered at the same time. The DME/P is capable of responding to both pulses in IA mode. Note that the 50% detection point for the Gaussian pulse is more than one microsecond later than the detection point for the COS/COS³ pulse. If the transponder applied the correct reply delay to the detection of the Gaussian pulse, this difference in detection points would move the reply to the Gaussian pulse to more than one microsecond later than the reply to the COS/COS³ pulse as measured at the interrogation and reply triggers. This faulty measurement technique would show that the DME/P did not meet reply delay accuracy requirements when in fact it does.

Value Averaging in Integer Measurement System: The second pitfall involves value averaging of integer measurements. Using any

kind of a clock to make time measurements forces integer values for measurements. When more resolution of a measurement is desired, the number of measurements used to determine the value can be increased and an average value taken. The average value computed must be at some higher resolution than the original measurement or the resolution improvement from averaging is lost. For example, if a reply pulse has a 5% to 30% time of 283.6 nanoseconds and the measurement clock is at 6.25 nanoseconds, in a 32 sample average there should be 20 measurements of 45 clocks and 12 measurements of 46 clocks. If the values are summed to 1452 clocks and an integer divide by 32 is completed, the result is 45 clocks at 6.25 nanoseconds which translates to 281.25 nanoseconds. If a floating point divide of the 1452 clocks by 32 is completed instead, the correct result is 45.375 clocks at 6.25 nanoseconds which translates to 283.59 nanoseconds. The same result is obtained by multiplying the 1452 clock sum by one thirty-second of the measurement clock period of 6.25 nanoseconds or 0.1953125 nanoseconds which results in a measurement of 283.59 nanoseconds. Either of the two methods for handling the increased resolution due to averaging can be used or other methods may be devised.

Detection Point Comparison - COS/COS³ Pulse vs. Gaussian Pulse

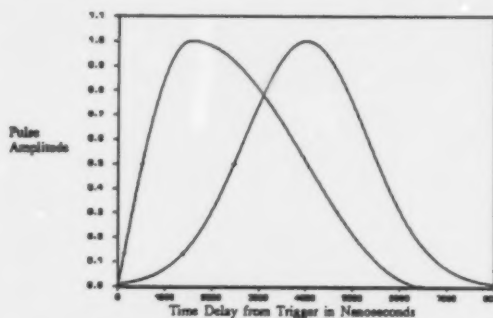


Figure 8

Test Equipment Calibration

The calibration of test equipment is extremely important in the determination and correction of any delay biases that exist in a DME/P system. The primary focus is to assure that the test fixture making the actual reply delay measurements is making as precise a measurement as is possible. This is done by first calibrating the reply delay measurement test fixture with equipment that has been calibrated to National Institute of Standards and Technology (NIST) (formerly NBS) standards.

Measurement Standards

The measurement standards used are for insuring the accuracy of the measurements to be made on the DME/P equipment.

Frequency Counter: A frequency counter is used to determine the current measurement clock frequency. The measured frequency is then used to determine a calibration factor that is used to adjust all reply delay measurements taken from the expected 80 MHz clock frequency to the actual clock frequency measured during calibration.

Digitizing Oscilloscope: A digitizing oscilloscope is used to make time domain measurements on the actual RF pulse being generated by the test fixture and a pulse for calibration. This oscilloscope provides sets of time and voltage data values corresponding to the envelope of the RF pulse being measured. The data gathered are used to determine frequency dependent delays that may exist in the test fixture circuitry.

Measurement Equipment Calibration

The measurement standard test equipment described above is used to calibrate the reply delay measurement test fixture. This calibration provides assurance that the measurement of reply delay is as precise as possible.

Measurement Clock Calibration: The fundamental measurement clock for the test equipment is 80 MHz. This is crystal controlled and therefore does not drift during measurements. There are, however, long term drifts that must be accounted for. This is done by measuring the actual frequency of the clock and then determining the factor that must be used to adjust the measured time delay value to the actual time delay value.

Reply Delay Measurement Offset Calibration: Besides the actual error in the time measurement that may exist, there is an offset error that exists in the points at which the measurements are taken. The transponder reply delay is measured based on the time the interrogation enters the transponder cabinet to the time the reply leaves the transponder cabinet. The test equipment actually measures the time delay from when the interrogation is formed in the test equipment until the reply is detected in the test equipment. These time differences must be determined as part of the calibration process and then removed from any reply delay measurement. This process must be completed to provide any assurance of the accuracy of the DME/P transponder when it is installed in the field.

Calibration of DME/P Equipment

After the test equipment has been calibrated, the DME/P transponder itself can then be calibrated. This calibration is completed to remove any biases that may exist in the system as configured and manufactured. When the DME/P transponder has been calibrated, the transponder is as accurate as possible for shipment from the factory. Precise adjustments based on site dependent parameters will maintain the accuracy of the DME/P transponder during site installation and use.

Basic Transponder Reply Delay Calibration

The basic transponder reply delay calibration consists of determining the bias error that exists in the system as factory built and then removing the bias through the programming of EEPROM (EAPROM) in the DME/P transponder. The bias error determination is done at a power level chosen to approximate the largest number of conditions that the transponder will see. A large number of reply delay measurements are made by the test equipment and the average bias error of the measurements is computed. The transponder calibration factor to offset the bias error is then calculated and programmed into the DME/P. Later factory tests verify that the calibration factors determined are correct.

DME/P Monitor Calibration

After the transponder has been calibrated, the monitor itself can be calibrated. This is done by taking a large number of sample outputs from the monitor filters and averaging the bias error reported by the monitor. The monitor calibration factor to center the monitor measurement is then calculated and programmed into the DME/P monitor. Later factory tests verify that the calibration factors determined are correct.

Reply Delay Performance Results

The proof of the validity of factory calibration supporting DME/P transponder field installation and usage comes from actual data taken in the field. The proof is presented in the form of tracking the reply delay performance of a DME/P transponder from the factory floor through final flight testing.

Transponder Performance Prior to Calibration and Test

The DME/P transponder is designed and built to operate over all environmental conditions within the requirements of any specification. Calibration of the DME/P is used to assure compliance with all requirements over all environmental conditions as well. Figures 9 and 10 show the FA and IA mode PFE performance for a DME/P directly off the manufacturing floor. Note that the performance is within the requirements and yet is not centered. Figures 11 and 12 show samples of the monitor measured FA and IA mode PFE performance for comparison to the test fixture measured performance. Note that the monitor is measuring PFE within limits yet slightly off from the test fixture. The monitor and test fixture comparison data are sampled due to the inability to obtain continuous data from the DME/P

monitor through the parameter reporting system.

FA Data Processing - PFE Performance Before Calibration

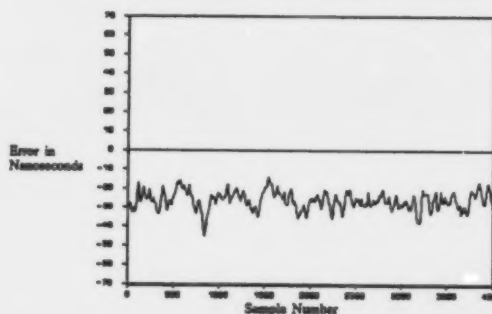


Figure 9

IA Data Processing - PFE Performance Before Calibration

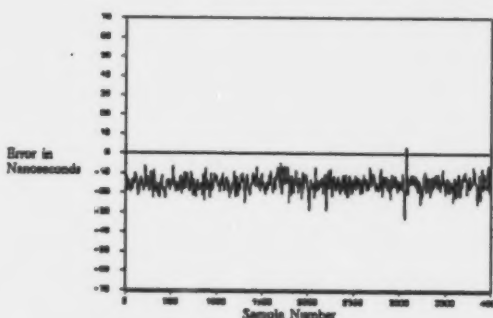


Figure 10

FA Data Processing - PFE Monitor Measurements Before Calibration

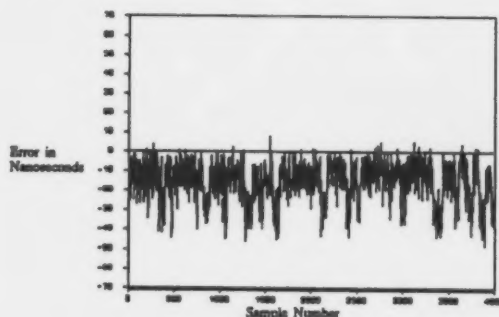


Figure 11

Transponder Performance After Calibration

Just prior to final acceptance testing, the DME/P is calibrated to correct the offset errors that exist in the system. Figures 13 and 14 show the FA and IA mode PFE performance of the transponder system after calibration. Note that the PFE is very nearly centered and provides safe margins for environmental drift in any time delay direction. Figures 15 and 16 show samples of the monitor measured PFE performance for comparison to the test fixture measured performance. Note that the comparison is now extremely close providing a high degree of assurance that the monitor reported values reflect actual system performance.

IA Data Processing - PFE Monitor Measurements Before Calibration

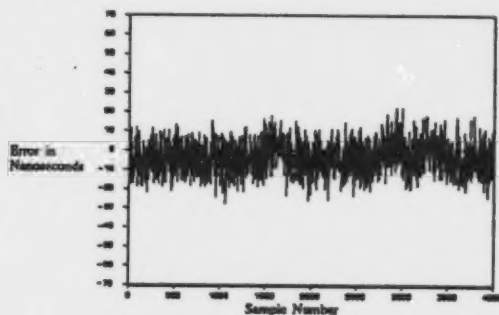


Figure 12

FA Data Processing - PFE Performance After Calibration

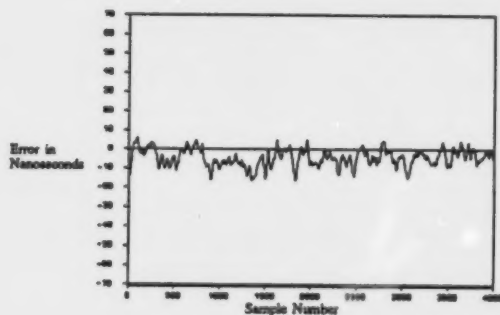


Figure 13

IA Data Processing - PFE Performance After Calibration

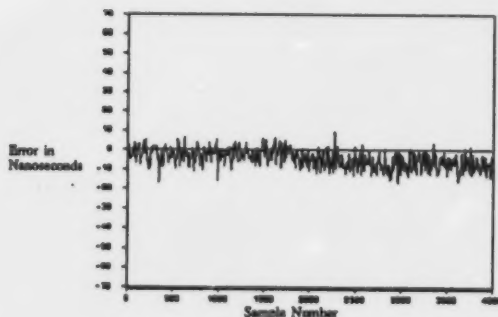


Figure 14

FA Data Processing - PFE Monitor Measurements After Calibration

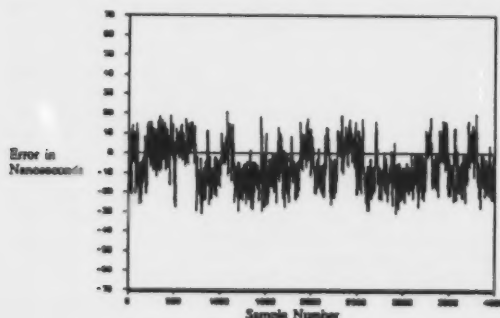


Figure 15

IA Data Processing - PFE Monitor Measurements After Calibration

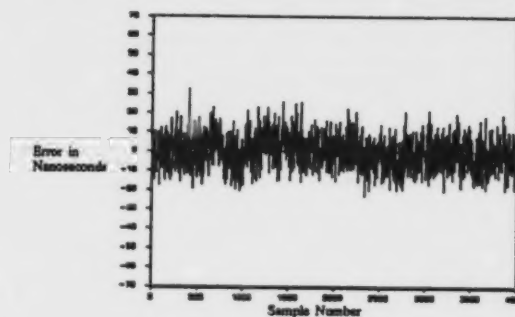


Figure 16

Transponder Performance at Site Installation

Upon completion of the site installation, stationary ground tests are completed to verify the accuracy of the installed system. The E-Systems DME/P transponder performance during stationary ground tests provided accuracy at or beyond the capability of the measurement equipment used for the ground testing.

Transponder Performance During Flight Tests

After the installation and ground tests are completed, the final checkout involves flight tests of the MLS system. Figure 17 shows complete DME/P system performance as measured during flight tests to commission an MLS system for use. Also shown in the figures are the FAA requirement limits for the complete system.

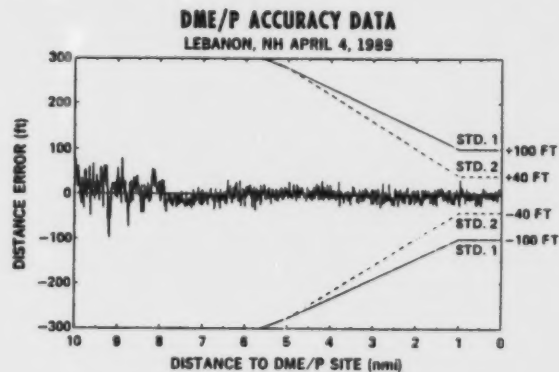


Figure 17

References

- [1] SC-149, Minimum Operational Performance Standards for Airborne Distance Measuring Equipment (DME) Operation Within the Radio Frequency Range of 960 - 1215 Megahertz, Radio Technical Commission for Aeronautics Document No. RTCA/DO-189, September 1985.
- [2] Annex 10, International Standards, Recommended Practices and Procedures for Air Navigation Services, Aeronautical Telecommunications, International Civil Aviation Organization, Volume 1, Fourth Edition, April 1985.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

TESTING MLS DME/P TRANSPONDERS AND INTERROGATORS

Paul E. Manning
Transportation Systems Center
Kendall Square
Cambridge, MA 02142

Eric J. Pesciotta and Joseph W. Ruggiero
Unisys Corporation
Transportation Systems Center
Kendall Square
Cambridge, MA 02142

Abstract: This paper describes the Microwave Landing System Precision Distance Measuring Equipment Automated Test System which was developed at the United States Department of Transportation, Transportation Systems Center. Included are descriptions of the automation software techniques and methods that are employed to overcome some instrumentation limitations. This paper assumes that the reader has some familiarity with the Precision Distance Measuring Equipment.

Introduction

The Microwave Landing System (MLS) is the internationally developed aircraft landing system that was approved by the International Civil Aviation Organization (ICAO) as a replacement for the Instrument Landing System (ILS). The MLS ground equipment transmits precise angle and range information which is processed by the aircraft MLS equipment to provide accurate positional information with respect to the runway threshold. The Precision Distance Measuring Equipment (DME/P) is a subsystem of the MLS that provides accurate range to runway threshold.

The Transportation Systems Center (TSC), an agency under the United States Department of Transportation, has been funded by the FAA since 1986 to develop a laboratory capability to independently assess the performance of DME/P ground transponders and airborne interrogators. To accomplish this task, TSC developed the DME/P Automated Test System (ATS). The ATS specifically addresses key paragraphs of the FAA specification for DME/P transponders [1] and the RTCA MOPS for DME/P interrogators [2]. The ATS is easily adaptable to ICAO Annex 10 standards.

Automated Test System Overview

One of the chief design requirements for the DME/P test system was that it be easy to use by a knowledgeable operator (i.e., an FAA technician). To this end, the ATS consists of software controlled test equipment and "cook book" procedures for performing the tests. It is important that the testing be automated to ensure that all units being tested are subjected to identical processes.

To automate the test system, physical links are established between the unit under test (UUT), the instruments, and the computer running the ATS software. The computer and instruments are connected via cables establishing an electrical link known as the Hewlett-Packard Interface Bus (HP-IB), which conforms to the IEEE-488 bus standard. The computer controls all instruments and transfers data over the HP-IB. The UUT and instruments are connected via coaxial cables, RF couplers, and attenuators. When establishing the RF link, care is taken not to exceed the maximum input power of the instruments.

To automatically measure the performance of the DME/P, the operator simply cables the unit under test to the ATS and uses the ATS menu-driven software. The testing is a two-step process: the first is Data Collection and the second is Data Analysis. Data

Collection is an interim process resulting in raw measurement data stored in fixed file formats on disk. This raw data is used by the Test Software in the Data Analysis mode to produce the final results of the measurement.

Test Equipment

Instruments for the ATS were selected on their adaptability to automation and their ability to meet DME/P performance requirements. To test DME/P for specification compliance, the ATS must have measurement accuracies that are better than those specified for the DME/P themselves. A goal in the development of the ATS was to achieve accuracies of an order of magnitude better than the DME/P. For example, in the Time Domain the partial pulse rise time must be 250 nanoseconds (nS) $\pm$ 50 nS, therefore, the ATS should measure this to within 5 nS.

Except for the Test Set and ARINC converters, the ATS is assembled entirely from Hewlett-Packard (HP) instruments. These instruments by themselves, however, do not satisfy all of the requirements for accuracies and automation. To compensate, the computer is needed in the system. The equipment in the ATS is as follows:

Test Set/Simulator.....Republic Electronics PDME/101A
Oscilloscope, Digitizing ..HP 54110D (1 GHz)
Spectrum Analyzer HP 8568B (100 Hz - 1.5 GHz)
Counter, Universal HP 5370B
Power Meter HP 438A
Computer HP Model 9000, Series 300
ARINC 568 to HP-IB TSC
ARINC 429 to HP-IB TSC

Since detailed information on the HP instruments is easily available, descriptions in this paper are limited to specific DME/P applications and instrumentation limitations that were discovered during the process of developing the DME/P test algorithms. More detailed information is provided for the Test Set/Simulator (TSS) since it is unique and information on it is not readily available to the reader.

Test Set/Simulator

The TSS is comprised of a basic unit with a separate display and keyboard. It has two basic operating modes: as an interrogator simulator (transponder evaluator) or as a transponder simulator (interrogator evaluator). In either operating mode, it can simulate and evaluate both DME/N and DME/P, with Peak Amplitude Find (PAF) and Delay Attenuate and Compare (DAC) detectors. The TSS can also measure and display peak power of received RF pulses, reply efficiency, mean reply delay time errors, pulse repetition frequency (PRF), and pulse spacing. Operation of the TSS can be controlled locally by its keyboard or remotely by an external computer connected to its HP-IB.

The TSS can operate in all DME/P channels with adjustable frequency offset and can measure reply efficiency for both DME/P and DME/N. It can generate COS-COS² pulses (DME/P) and Gaussian pulses (DME/N). It can simulate reply efficiency, range, velocity, acceleration, multipath (delay, amplitude, phase and phase rate), squitter and traffic loading. Pulse pair spacing, RF power and frequency can be varied and an independent auxiliary channel is also provided which can simulate adjacent channel operation.

The TSS contains Path Following Error (PFE) and Control Motion Noise (CMN) filters which estimate UUT accuracy. The front panel of the TSS has Start/Stop connectors that are wired to triggers from the PAF/DAC detectors. These output triggers represent the interrogation (Start) and reply (Stop) as detected by the TSS and, when connected to the HP 5370B Universal Counter, provide a measurement of reply delay variations. The ATS software uses this feature as opposed to the internal TSS filters to precisely compute UUT error.

Parallel channels and dual modulators allow for simultaneous generation of Main and Echo (Multipath) pulses. The RF pulses are controlled by precision clocks, phase shifters, and modulators. A frequency synthesizer with an accuracy of 0.001% is used to provide the RF carrier to drive the Main and Echo channels. When testing requires two different frequencies (e.g., adjacent channel interference), a second synthesizer independently drives the auxiliary channel. In addition to the I/O port, two other RF ports are provided on the front panel for use in calibrating the TSS and to monitor pulse performance.

Two pilot pulse counters, one for DAC triggers and one for PAF triggers, are used to adjust the zero mile delay counters to correct long-term drifts. They measure the total RF delay which is input to the CPU which statistically analyzes samples and supplies corrections for the zero mile delay counters. During calibration of the TSS it was found that the TSS maintained internal delays to ± 3 nS, with no detectable changes with frequency or power output variations.

The TSS provides the precision and functions necessary for testing some DME/P specification parameters, such as multipath and traffic loading, but it is an expensive instrument. Some ICAO members have expressed the opinion that it would be desirable that a less costly instrument be developed or some other means for performing these tests be suggested. The DME/P is capable of measuring time differences in units of nanoseconds, and any evaluation instruments must be at least as precise.

Test Software

The software provides for instrument control, data collection, storage, analysis, and reports. All measurements are performed in a controlled, directed manner. The primary programming language is HP Rocky Mountain BASIC, which provides a structured programming language with all modern programming constructs as well as a very close relationship with the operating system and hardware. However, the selection of BASIC traded off execution speed for programmer convenience. To compensate for time-consuming processes, such as curve fitting and digital filtering, compiled Pascal functions are linked to the software.

The ATS Test Software automates performance testing with a user-friendly interface. This interface was designed to simplify the actions and minimize the keyboard entry required to perform tests. Also, testing with the software requires only a limited knowledge of computers. To avoid operator confusion and to expedite testing, screen layout, keyboard assignment, menu processing, and file handling standards are well defined and consistent.

The test software program consists of organized libraries that logically flow from the user interface and test components. Low-level libraries perform basic tasks such as screen formatting, data entry, menu processing, and datafile management. For each instrument used in the ATS, a library performs tasks specific to the instrument's function. At the highest level are libraries that perform data analysis of time domain, frequency domain, and accuracy.

Data Collection

In the Data Collection step, the ATS software is used to control the operation of the digitizing instruments. To collect data the software will set up the instrument, as prescribed by TSC staff, and command it to digitize the UUT's RF signals. The software then transfers the digital data from the instrument to a computer disk datafile. In general, all data is stored on disk in its unformatted binary form. This ensures data integrity and conserves disk space. These datafiles contain all the information required for the Data Analysis step.

Data Analysis

In the Data Analysis step, the ATS software is used to process the raw datafiles created by the Data Collection step. The Data Analysis generally adds a level of improved accuracy over raw data alone. The improved accuracies result from mathematical manipulation of the raw data, at times combining processed data from multiple instruments. Data Analysis may be performed in a batch processing mode, printing reports automatically and providing measurement results with minimal operator intervention. All reports contain historical information about the test and UUT, as well as the actual measurement results. In many cases tabular results are available, indicating specification pass/fail status.

Time Domain Tests

To perform Time Domain Data Analysis, data is collected from the HP 54110D Digitizing Oscilloscope (DO). The DO is commanded to acquire 19 digitized waveforms, each capturing a distinct segment of the RF pulse modulated envelope. These 19 digitizations contain comprehensive data on both the first and second pulse of the pulse pair. All 19 are necessary to test for compliance with the FAA specifications.

Collecting data from the DO is a two-step process. To allow for variation in UUT pulse shapes, data collection is accomplished with operator interaction, creating a datafile containing 19 setup records. Each record holds values for Volts/Div, Sweep Speed, Vertical Offset, and Time Base Delay. To help create the setup datafile, the software will initialize the oscilloscope to default values for each of the 19 records, allow operator adjustments, and store the resulting DO settings.

The second step performs the actual data collection without operator intervention. To do so, the software reads the "setup datafile" for proper settings, digitizes data on the oscilloscope, and stores 19 digitized waveforms in a separate "pulse datafile." Under normal conditions, one DO setup datafile may be used to collect many pulse datafiles. However, to account for the effects that frequency may have on the pulse, a unique setup datafile must be created for each DME channel tested.

The precision of Time Domain measurements varies with pulse slope, because all measurements involve the intersection of time and voltage, specified as a percentage of the pulse peak. Thus, the DO resolution on both axes, voltage and time, becomes critical. Early attempts to acquire high resolution digitizations, necessary to achieve the desired precision, uncovered some limitations of the DO. To achieve the best resolution, the vertical scale (voltage) and offset of the DO are adjusted to fully display only the desired voltage frame of the positive half of the pulse. In addition, the horizontal scale (time) is adjusted such that only the desired time frame is displayed. At this high resolution, the DO shows the results of digitized noise, which makes measurements dramatically less precise. This noise results from the combination of over-driven Charged Coupled Devices, over-driven D/A converters, and the increased gain of the higher resolution. Thus, for optimal resolution, vertical scaling is limited. In addition, due to the negative side of the pulse, the offset is reduced to no more than one half the vertical scale. Effectively, the accuracy of Time Domain Measurements is directly related to vertical quantization error and pulse slope.

Quantization error relates to voltage levels. On the DO (due to the above restrictions) there are a maximum of 64, which allows errors up to $\pm 1.6\%$ of the pulse peak voltage. This resolution combines with horizontal resolution to achieve time measurement accuracies on the DME/P pulse shape of ± 16 nS. This does not meet the requirements for Time Domain testing. To further improve the accuracy, measurements are not made on raw DO data, but on polynomials derived from the data. For simple measurements, the polynomial is a result of a least squares curve fit on at least 20 points of the raw DO data. For more critical measurements, up to 500 data points may be used. It has been shown that the additional resolution picked up by curve fitting results in Time Domain measurement accuracies to within 3 nS.

All measurements are made relative to the peak of the pulse. Therefore, accurate determination of the pulse peak is essential. To do so, many algorithms were examined, ranging from one as simple as maximum value to one involving a complex histogram technique. Because of the possibilities of noise, most were flawed to one degree or another. As a result, the curve fitting algorithm proved to achieve the most consistently accurate peak measurements. The pulse peak is determined by estimating the peak using the maximum voltage value in the digitized pulse peak raw data, taking 95% of this value, finding the corresponding raw time locations on both sides of the pulse, and fitting a polynomial to the digitized data between these points. The zero slope of the polynomial then determines the processed peak voltage.

Time domain tests include simple measurements involving the determination of time between two points on the pulse, and more sophisticated measurements involving all points on an entire portion of the

pulse. Simple measurements include pulse width (50% on the leading to 50% on the trailing edge), full rise time (10% to 90% leading edge), and pulse spacing (50% leading edge first to second pulse). For each of these values times are estimated using the processed peak voltage and the digitized raw data, 20 points of raw data surrounding these estimates are used for curve fits and the resulting polynomials are solved for time using the desired voltages. More complex testing is performed to ensure linearity in the partial rise time region (5% to 30%) on the leading edge of the pulse. Linearity is determined by fitting a third degree polynomial between the 4% and 31% points, solving this polynomial for the 5 and 30% points, computing the instantaneous slope (i.e., first derivative of the polynomial) at 1% increments from 5% to 30%, and comparing the results to the straight line slope. All ATS accuracy measurements are based on Virtual Origin (VO): the intersection of the baseline with a straight line drawn from the leading edge 5% and 30% points.

Frequency Domain Tests

The HP 8568B Spectrum Analyzer (SA) and the DO are used for Frequency Domain evaluation of both DME/P transponders and interrogators and to detect inband spurious noise in transponders. The DO is used to determine the effective square pulse width for the sideband and peak power measurements.

The SA operates in the 100 Hz to 1500 MHz frequency range, has its own microprocessor and is controlled by and provides data over the HP-IB interface with the computer. The harmonics test and out-of-band spurious noise test require a higher frequency range and another spectrum analyzer is used in the ATS for frequencies greater than 1500 MHz (another spectrum analyzer is now available from HP that will cover the entire frequency range required by DME/P specifications).

Testing with the SA for both transponders and interrogators takes place in two steps with two datafiles being generated. RF stability, pulse spectrum sidebands, and peak power are determined in a batch mode with a single pulse train and the results of these tests are placed in one datafile. These tests can be performed with a double pulse train, however, the results of the sideband power measurement are not conclusive. The RF stability test is performed with this group of tests because the results are used in the sideband and peak power measurements. The second group of tests, also performed in a batch mode, include spurious noise and energy band tests. These are performed on the dual pulse train which is used in the standard operation of the DME/P.

The bandwidth of the SA receiver can be adjusted using the resolution and video bandwidths. The IF bandwidth is controlled by the resolution bandwidth. The video bandwidth controls the analyzer's post detection filter and is set at least as wide as the resolution bandwidth. A line spectrum is produced when the IF bandwidth is less than the PRF and a pulse spectrum response is produced when the IF bandwidth is greater than the PRF. For DME/P testing, the IF bandwidth must be greater than 1.7 times the PRF and the scan time must be equal to or greater than 10/PRF. For more details on DME/P spectrum analysis, see reference [3].

Test Set Calibration

The TSS has a front panel RF port (Port B) which is internally coupled to the RF input of the internal PAF/DAC detector. The interrogation and reply RF pulses can be observed by connecting the DO to Port B. Since True Delay of a unit under test is defined by DME/P specifications as the delay from interrogation virtual origin (VO) to reply VO, ATS software can utilize the DO to determine the time at these virtual origins and thus determine True Delay at the TSS detector.

The TSS outputs a very stable RF pulse that is well within DME/P specifications (particularly the critical 5% to 30% leading edge linearity) and is a perfect source for measuring delay through all cables, couplers, and attenuators that may be used to couple the TSS with a unit under test. All these delay measurements are made by measuring VO time.

To calibrate the TSS, it is connected via cables and a coupler interface to an interrogator which has stable interrogations with pulse shape conforming to DME/P specifications. The ATS and the DO are used to measure interrogation to reply delay out of the coupler interface. This delay is adjusted by the delays through the interface and cables to arrive at the TSS true front panel delay. The difference between this delay and the delay measured at the TSS Port B can be used as a calibration offset. When other interrogators or transponders are tested, the delay can be conveniently measured at TSS Port B and the offset used to determine the true delay at the TSS front panel and back through the connecting cable to the unit under test.

Interrogator Accuracy Tests

VO true delay is measured at Port B of the TSS. Adjusting this delay with the TSS calibration offset plus the delays through the interconnecting cables and attenuators results in the true delay at the UUT. This measurement is made for both initial and final approach modes.

The ATS software is used to vary the TSS settings so that an interrogator can be evaluated for accuracy under conditions with and without interference: error vs. receiver input power, range, reply efficiency, and multipath. At each setting, the ATS is used to simultaneously collect data from the Counter connected to the TSS Start/Stop connectors (TSS measured delay variations) and from an ARINC converter connected to the interrogator under test (UUT measured delay).

The ATS software uses the following algorithm to determine UUT accuracy corresponding to each TSS simulated condition. The TSS measured delay variations are summed with the true VO delay at the UUT. These values (true VO delay with variations) are passed through a 10 RAD low pass filter. These filtered true delay values are then compared to the UUT measured delay. The differences between the two are passed through the ATS software PFE and CMN filters and the resulting errors are reported.

For TSS reply efficiencies less than 100%, there is a problem with the Counter missing valid replies. After receiving a start trigger the counter will ignore all subsequent start triggers until a stop trigger is received. Missed replies will appear as high counts and one or more valid delays will be missing from the data. This limitation can be

overcome by using two counters, or as implemented in the ATS, by using a circuit that automatically generates a stop trigger after a specified time if a stop trigger does not occur. The ATS software is then used to filter out these missed replies.

Transponder Accuracy Tests

A similar method is used to test Transponder accuracy using the TSS to simulate various conditions with and without interference.

The ATS is again used to measure the VO True delay at the TSS Port B. This delay is adjusted by the TSS calibration offset, plus the delays through the interconnecting cables and attenuators, and results in the true delay at the transponder under test. The ATS is used to collect data from the Counter connected to the TSS front panel Start/Stop connector. These TSS measured delay variations are summed with the UUT true delay. These values (true delay with variations) are then compared to the specification delay. The differences between the two are passed through PFE and CMN filters, and the resulting PFE and CMN errors are reported.

Conclusions

The TSC DME/P Automated Test System has been demonstrated to be capable of rapidly and accurately assessing the performance of both DME/P transponders and interrogators. It has been used to assess the performance of DME/P transponders (E-Systems and Sierra-Nevada), DME/P interrogators (Plessey, SEL) and DME/N interrogators (King, Honeywell), and its utility has been confirmed, in some aspects, by the acceptance of the software used in the ATS by other governmental agencies and industrial firms, both in the United States and abroad.

References

- [1] FAA Specification for Precision Distance Measuring Equipment, FAA-E-2721/3b.
- [2] RTCA Minimum Performance Standards (MOPS), DO-189, September 1985.
- [3] R. Valone, R.G. Baldwin and K.A. Tench, "MLS DME/P Pulse Spectrum Analysis," Project Memorandum DOT-TSC-FA895-PM-87-18, U.S.D.O.T., Transportation Systems Center, Cambridge, MA, November 1987.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

VALIDATION TESTING OF DME/P GROUND
TRANSPONDERS AND AIRBORNE INTERROGATORS

John F. Michaels
Republic Electronics Co.
200 Wireless Blvd.
Hauppauge, NY 11788

Abstract

In order to properly validate equipment performance in its intended environment, the test equipment must be capable of simulating the environment over its dynamic range of expected variations up to and including projected worse case conditions. This paper describes the required performance and implementation for validation testing of DME-P Ground Transponders and Airborne Interrogators.

RTCA/DO-189 Minimum Operational Performance Standard (MOPS) for Airborne DME is the baseline for interrogator testing, while FAA-E-2721/3b is the baseline for transponder testing. Numerous other documents were used along with special committee participation and development activities to contribute to the information contained herein.

DME-P system accuracies are specified in terms of Path Following Error (PFE) and Control Motion Noise (CMN) relating the signal to aircraft position errors and flight control dynamics. Factors contributing to degrade system accuracies are interrogator and transponder efficiencies, downlink and uplink garble and multipath. DME error and power budgets must be properly allocated to each system component. Because of the importance of the implemented measurement technique, both PFE and CMN separate and composite measurement configurations are discussed, including contribution to range error and an indication when the error measurement exceeds 5% within the defined sliding window.

The performance of a test set simulator (TSS) will be described in two parts: as an interrogator evaluator and as a transponder evaluator.

As a transponder evaluator, operation in a traffic loading mode is required in addition to evaluating performance. The criteria for PFN and CMN measurements described also apply here. As part of traffic loading tests, random interleaved IA and FA pulses in adjustable proportion (0% IA and 100% FA to 100% IA and 0% FA) with intermediate combination need to be provided. The characteristics of pulses are precisely controlled in amplitude, time, phase and frequency.

Finally, for validation, TSS accuracies need to be an order of magnitude better than that required by the system and at least three times better than each component. The special techniques required for calibration are described.

Introduction

Since the beginning of 1970, Republic Electronics Co. (REC) has been involved with TACAN simulation. This included the development of a simulator early in 1970 for Standard Elektrik Lorenz (SEL) with a .01 NM range accuracy. SEL in the 70's, successfully developed a tactical landing system based on TACAN called Sector TACAN (SETAC), and REC was awarded a contract to develop a SETAC simulator for airborne equipment testing. It was at this time that REC evolved their first microprocessor controlled simulator and closed loop dual channel RF modulator.

Since 1981, ground and airborne DME/P equipments have been under development at SEL.¹ As a result, REC was awarded a contract in 1981 for a DME/P simulator capable of outputting both standard DME/N and DME/P cos/cos² pulses within the two pulse/two mode system defined by the International Civil Aviation Organiza-

tion (ICAO).^{1,2,3} This equipment featured a delay accuracy of better than 5 meters and outputted range replies with superimposed echoes.¹ It was the result of its use, and data obtained by SEL, that REC was successful in receiving a contract in early 1985 from the Dept. of Transportation/Transportation System Center (DOT/TSC) for two units for what is known as MLS DME N/P Test Set/Simulator (TSS), Model PDME-101A that is described herein. The delay error specified for the DME/P FA test mode was not to exceed ± 20 nsec at the output of the PFE filter and ± 18 nsec at the output of the CMN filter.² The delivered units exceeded both requirements.

The use of the TSS by DOT/TSC as a transfer standard for transponder and interrogator testing is also discussed in these proceedings by P. Manning in his paper, "Testing MLS DME/P Transponders and Interrogators".

TSS Description⁶

The TSS is comprised of the basic unit with a separate display and a keyboard. It has two basic operating modes: as an interrogator simulator (transponder evaluator) or as a transponder simulator (interrogator evaluator). In either operating mode, it can simulate and evaluate both DME/N and DME/P, with PAF and DAC detectors. The TSS also can measure and display peak power of received RF pulses, reply efficiency, mean reply delay time errors, PRF and pulse spacing. The TSS is controlled via a keyboard or an IEEE-488 bus and can output data to either an RS-232 line or IEEE bus. Operation of the TSS can be controlled locally by the keyboard or remotely by an external computer.

TSS Functions and Specifications

The TSS can operate in all DME/P channels with adjustable frequency offset and can measure accuracy (PFE and CMN) and reply efficiency for both DME/P and DME/N. It can generate COS-COS² pulse (DME/P) and gaussian pulses (DME/N). It can simulate reply efficiency, range, velocity, acceleration, multi-path (delay, amplitude, phase and phase rate), squitter and traffic loading. Pulse pair spacing, RF power and frequency can be varied and an independent auxiliary channel is also provided which can simulate adjacent channel operation. Figure 1 shows the TSS display when it is in a transponder simulator mode and Figure 2 shows the TSS display when it is in an interrogator simulator mode. All parameters are controllable in local operation. The detailed specification for the TSS is described in Reference 7.

Interrogator Validation

Figure 3 is a simplified block diagram of the TSS as an airborne interrogator tester.⁷ The TSS has the capability to simulate range and range dynamics in addition to simulating the appropriate transponder delay dependent upon the respective interrogation mode and selected operating channel. The operator inputs the desired range and range dynamics via the keyboard and the displayed range is part of the total delay but excludes the TSS simulated transponder delay (zero

mile delay). The range processed by the Interrogator Under Test (IUT) is displayed. (See Figure 1.)

The PFE and CMN 95-percentile range errors displayed are determined by using the difference between the range data obtained from the IUT and the TSS instantaneous reference range (see Fig. 3). In addition, separate Go/No Go indicators are provided when more than 5% of the PFE and/or CMN measurements exceed the error tolerance of the IUT.

The TSS determines the PFE and CMN range errors of the IUT as follows (see Fig. 3):

1) PFE Error - The range output from the IUT is compared to the Instantaneous Reference of the TSS. For a 10 second measurement interval, this error data is passed through a low pass filter having a 1.5 rad/sec (0.24 Hz) cut-off frequency with a roll-off characteristic approaching 12 dB/octave. The output of this filter is used to determine the 95-percentile value which is indicated on the PFE display. The reading on this display is fixed until updated at the completion of the next measurement period, which must occur 1 second later in time (1 second update interval for a 10 second sliding window measurement period).

2) CMN Error - Over the same 10 second measurement interval described above for the PFE error, the same compared data (error) is measured at the output of either a high-pass or a band-pass filter, selectable from the keyboard or the IEEE-488 bus. The high-pass filter has a (1) cut-off frequency of 0.5 radians/sec (0.08 Hz), (2) an essentially flat response above its cut-off frequency to at least 10 radians/sec, and (3) a roll-off characteristic approaching 6 dB/octave. The band-pass filter has (1) a lower cut-off frequency of 0.5 radians/sec, (2) an upper cutoff frequency of 10 radian/sec (1.6 Hz), (3) essentially flat response between its cutoff frequencies, and (4) a rolloff characteristic approaching 6 dB/octave. The output of these filters is used to determine the 95-percentile value of the data provided by the interrogator and indicated on the CMN display. Using the high-pass filter, the CMN bandpass characteristic for the interrogator measurement is a composite of the interrogator low-pass filter and the TSS high-pass filter. The reading on the TSS display is fixed until updated at the completion of the next measurement period which must occur 1 second later in time (1 second update interval for a 10 second sliding window measurement period).

The sliding window measurement interval described above is automatically selected by the TSS at 10 seconds (FA-mode) or 40 seconds (IA-mode).

Over flight can also be simulated. The slant range rate can be set to any value between 0 and 1000 kts inbound, in 1 kt increments, starting at a preprogrammed range and continuing until a minimum range of 0 to 5 nmi (adjustable by keyboard control) is achieved. The sign of the range rate will then reverse and continue outbound to the maximum range.

The TSS reply efficiency of the DIRECT channel pulse pairs is adjustable over the range of 0 to 100 percent in either the FA or IA-modes. Accordingly, the ECHO channel will only generate ECHO pulse pair replies in synch with those pairs in the DIRECT channel.

The output pulse pair signal level from the DIRECT and the ECHO channels is independently controlled from -10 dBm to -120 dBm. In addition, the pulse code spacings of the DIRECT and ECHO channels is independently controlled. The squitter pulse pairs in each channel has the same pulse code spacing as the DIRECT and ECHO pulse pairs, respectively.

Echo signals (multipath) are provided for all pulses output by the DIRECT channel, with independent control provided for RF level, reply delay, squitter/ID delay, phase and phase rate (RF scalloping).

The squitter rates are independently controlled and adjusted in each channel with a non-uniform distribution in conformance with FAA E-2721/3b⁶, and approximately that of MIL-STD-291B at 2700 FPS. ID pulses are also provided in conformance to FAA E-2721/3b.⁶ Pulse threshold selection is provided with these options:

(a) "PAF": Interrogations and reply pulses are triggered at the 50% (6 dB) leading edge simulating the conditions specified in FAA E-2721/3b.

(b) "DAC": Interrogations and reply pulses are triggered using a DAC whose parameters are 100 nsec and 6 dB attenuation.

(c) "AUTO": PAF or DAC thresholding is automatically applied dependent on N or P pulse spacings.

In addition to the above described performance as an interrogator tester, the TSS also functions as a transponder simulator capable of replying to two interrogators and properly responding to FA and/or IA-mode interrogations.

Transponder Validation

Figure 4 is a simplified block diagram of the TSS as a transponder tester.⁷ Since many of the test set/simulator features are identical to the interrogator tester configuration, only the areas of difference will be described in this section. The transponder under test is designated TUT.

The TSS has the capability to measure the transponder delay of the TUT. In addition, the appropriate reference transponder delay, used for comparison with the TSS measurement is only dependent upon the respective interrogation mode (IA or FA) and the selected operating channels; known "a priori" by the TSS as operator inputs via the keyboard.

An offset adjustment is provided by the TSS to compensate for the RF interconnection cable between the transponder and the antenna present during normal operation or between the transponder and the TSS during testing. When using a non-zero offset value, the PFE error displayed by the PDME-101A can have an additional uncertainty of ± 3 nsec due to the offset quantization error. This value is displayed. The TSS provides the following delay displays when testing the TUT:

1) The TSS Reference Transponder Delay (RTD) - This is the zero mile fixed reference delay of the TSS simulating the FA or IA-modes with X, Y, W and Z pulse code spacings. The RTD's entered into the TSS by the operator via the keyboard control to determine actual transponder error.

2) The measured transponder delay actually processed and outputted by the TSS, updated once per second.

The PFE and CMN 95-percentile transponder delay errors are displayed by the TSS as the difference between the measured transponder display and the zero mile delay. In addition, separate PFE and CMN GO/NO GO indicators are provided to indicate when more than 5 percent of the PFE and/or CMN measurements exceed the error tolerance of the TUT as previously described.

The TSS determines the PFE and CMN transponder delay errors of the TUT using two different and distinct measurement standards. Either of the standards are selectable from the keyboard or via the IEEE-488 Bus control.

The TSS can determine the PFE and CMN transponder delay errors of the TUT predicated on a 95-percentile basis as stipulated in the following:

1) PFE - The transponder delay provided by the TUT and measured by the TSS is compared to the RTD. The result of this comparison is the error associated

with the TUT.

2) CMN Error - Over the same 10 second measurement interval described above for the PFE error, the same compared data (error) is measured at the output of either a high-pass or a band-pass filter, selectable from the keyboard or the IEEE-488 bus.

In addition to the 95-percentiles test standard delineated in this specification, the TSS can determine the transponder integrity parameters of the TUT predicated on the standards outlined in Table 3-3 of FAA E-2721/3b and the Test Set Simulator Specification.

In order that the TSS meets specified accuracy requirements, the TUT reply pulse peak power at the input to the TSS shall be within 100 watts ± 3 dB.

The TSS, when configured to test the TUT, has the capability to select two interrogation modes/rates: FA-mode with a PRF of 40 PPS and IA-mode with a PRF of 16 PPS. The mode is selected via the keyboard and displayed. The interrogation rate has a ± 6 milliseconds RMS jitter with uniform distribution.

When the TSS acts as an interrogator, it provides an additional failure detection mode for Go/No Go pass/fail indication evaluating reply data compared to a reference tolerance inputted via the keyboard. In the catastrophic mode, the following applies: The TSS detects and indicates an out-of-spec or catastrophic transponder failure when more than 5% of the errors exceed the input reference. The assessment will be made on a pulse-to-pulse measurement at the output of a 10 radian per second filter (FA or IA selectable). The FA mode filter is implemented using the following relationship.

$$FA_{LP_n} = 0.1111 * (FA_{X_n} + FA_{X_{n-1}}) + 0.7777 * FA_{LP_{n-1}}$$

where: FA_{LP_n} = Output of low pass 10 radian/second filter

$FA_{LP_{n-1}}$ = Previous output of filter

FA_{X_n} = Difference between current reply delay reading and nominal

$FA_{X_{n-1}}$ = Difference between previous valid reply delay reading and nominal

The IA mode filter has been implemented using the following relationship.

$$IA_{LP_n} = 0.2369 * (IA_{X_n} + IA_{X_{n-1}}) + 0.5213 IA_{LP_{n-1}}$$

where: IA_{LP_n} = Output of 10 radian/second filter

$IA_{LP_{n-1}}$ = Previous output of 10 radian/second filter

IA_{X_n} = Difference between current reply delay reading and nominal

$IA_{X_{n-1}}$ = Difference between previous valid reply delay reading and nominal

When a failure is detected, the Go/No Go indicator displays a fail condition on the terminal screen and further data is rejected after an additional 350 data points have been taken. Errors up to 250 nanoseconds are quantified. Those above 250 nanoseconds are not quantified but will be labelled as failed. The TSS maintains in memory the last 650 data points before failure and 350 data points after failure. Transfer of the 1000 data points to an external computer may be made by request via the IEEE-488 bus. Normal PFE and CMN measurements and calculations are suspended during the failure detection operation which is operable for both FA and IA modes, i.e., 40 and 60 Hz, respectively.

The TSS is capable of simulating on channel random interleaved IA and FA traffic loading interrogations. The IA and FA interrogations are adjustable from zero percent IA and 100 percent FA to 100 percent IA and zero percent FA.

The TSS is also capable of providing off channel (pulse code and/or frequency different than that of the interrogation signal) traffic loading simulation. The second VCO is used to simulate off channel signals

at frequencies other than that of the interrogation frequency. The pulse code spacing of the DIRECT and ECHO channel signals are independently controlled via keyboard entry.

Test Set/Simulator Calibration

The test set calibration procedure was evolved by DOT/TSC.⁴ Various calibration ports are provided by the PDME-101A to simplify calibration. Stated in reference 4: "Delays through the TSS RF modules have been found to be extremely stable over frequency, power levels, time and temperature. Only if the RF modules or internal RF cables are replaced would the user have to repeat the entire calibration procedure."

The TSS design includes a set of internal switches that can be set to adjust internal zero mile delays when the TSS is acting as a transponder.

References

1. K. Becker, A. Muller, H. Vogel, "Precision Distance Measuring Equipment for the Microwave Landing System", *Electrical Communication*, Vol. 58, 1984, Standard Elektrik Lorenz AG, Stuttgart, Federal Republic of Germany.
2. R. J. Kelly, Member IEEE, Bendix Corp., "System Consideration for the New DME/P International Standard", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-20, No. 1, Jan. 1984.
3. R. J. Kelly and D. R. Cusick, "Distance Measuring Equipment and Its Evolving Role In Aviation", *Advances in Electronics and Electron Physics*, Vol. 68.
4. P. Manning, J. Ruggiero, E. Pesciotta, R. McCarthy, "Description of the TSC MLS DME/P Automated Test System", DOT/Transportation System Center (TSC), with contribution by R. Vallone, FAA.
5. "Minimum Operational Performance Standards for Airborne Distance Measuring Equipment (DME) Operation Within the Radio Frequency Range of 960-1215 Megahertz", RTCA/DO-189.
6. "Microwave Landing System Ground Equipment Precision Distance Measuring Equipment (DME/P)", Specification, FAA-E-2721/3b, Oct. 27, 1983.
7. Republic Electronics Co. and Department of Transportation/Transportation System Center DOT/TSC, "Test Set Simulator Specification", Nov. 10, 1987.

TEST SET INTERROGATOR TESTER/TRANSPONDER SIMULATOR		
DIRECT CHANNEL	UNIT UNDER TEST	
MODE---(M,X,Y,Z) X	RNG----- (NM) 0.000	INT NO OUT--- (NM) I
ONE PAGE---(P,N) P	RNG RATE---(RTS) 0.	PFE ERROR---(NS) I
CHNL NO----- 1.	RNG ACC-(FT/S/S) 0.	PFE REF TOL-(NS) 67.
FREQ OFFSET(Y,N) N	OVERFLIGHT-(Y,N) N	PFE---(PASS,FAIL) I
CHNL FREQ + OFF- 981.992	RIN RNG--- (NM) NA	OWN ERROR---(NS) I
REPLY EFF---(Z) 70.	OUTPUT---(P,C) P	OWN REF TOL-(NS) 67.
APPROACH---(I,F,N) A	PULSE SPACE-(US) 12.0	OWN---(PASS,FAIL) I
ID TONE---(Y,N,K) K	SQUIT RATE-(PPS) 2700.	RF PK PWR-(WATT) 0.
ID + REPLY-(Y,N) N	PULSE SHP(C,G,I) C	PULSE SPACE-(US) 0.00
	RF LEVEL---(DBM) -50.	INT PPS---(PPS) 0.0
ECHO/AUX CHANNEL		
ECHO/AUX---(E,A) E	RF LEVEL---(DBM) -50.	ID/SQUITTER(I,S) NA
CODE---(M,X,Y,Z) NA	OUTPUT---(P,C) P	DELAY---(US) 0.000
CHNL NO----- NA	PULSE SPACE-(US) 12.0	PHASE---(DEG) 0.00
FREQ OFFSET(Y,N) NA	SQUIT RATE-(PPS) NA	PHASE RATE---(HZ) 0.0
CHNL FREQ + OFF- NA	PULSE SHP(C,G,I) NA	

Arrows select field, Letters, Digits edit, (Return) accept number
(F) continue, (P) print screen

MENUS AB

Figure 1.

TEST SET TRANSPONDER TESTER/INTERROGATOR SIMULATOR		
DIRECT CHANNEL	ECHO/AUX CHANNEL	UNIT UNDER TEST
APPROACH---(I,F,N) F	ECHO/AUX---(E,A) E	REF TRANSP DEL(US) 50.000
CODE---(M,X,Y,Z) X	DELAY---(US) 0.000	ACT TRANSP DEL(US) I
CHNL NO----- 1.	PHASE---(DEG) 0.00	PFE ERROR---(NS) I
FREQ OFFSET(Y,N) N	PHASE RATE---(HZ) 0.0	PFE REF TOL-(NS) 67.
CHNL FREQ + OFF- 1024.990	APPROACH---(I,F) NA	PFE---(PASS,FAIL) I
RF LEVEL---(DBM) -50.	CODE---(M,X,Y,Z) NA	OWN ERROR---(NS) I
PULSE SHP(C,G,I) C	CHNL NO----- NA	OWN REF TOL-(NS) 67.
PULSE SPACE-(US) 18.0	FREQ OFFSET(Y,N) NA	OWN---(PASS,FAIL) I
INTER---(F,J,P) F	CHNL FREQ + OFF- NA	RF PK PWR-(WATT) 0.
INTER RATE-(PPS) 40.	INTER---(F,J,P) NA	PULSE SPACE-(US) 0.00
OUTPUT---(P,C) P	INTER RATE-(PPS) NA	TRPO PPS---(PPS) 0.0
TRAP LOAD---(Y,N) N	RF LEVEL---(DBM) -50.	REPLY EFF---(Z) I
Z FA/(A)----- NA	PULSE SHP(C,G,I) NA	OWN FILTER---(B,N) N
CABLE OFFSET(NS) 0.	PULSE SPACE-(US) 18.0	PFE FILTER---(M,N) N
	OUTPUT---(P,C) P	CRY CODE---(Y,N) N

Arrows select field, Letters, Digits edit, (Return) accept number
(F) continue, (P) print screen

MENUS AB

Figure 2.

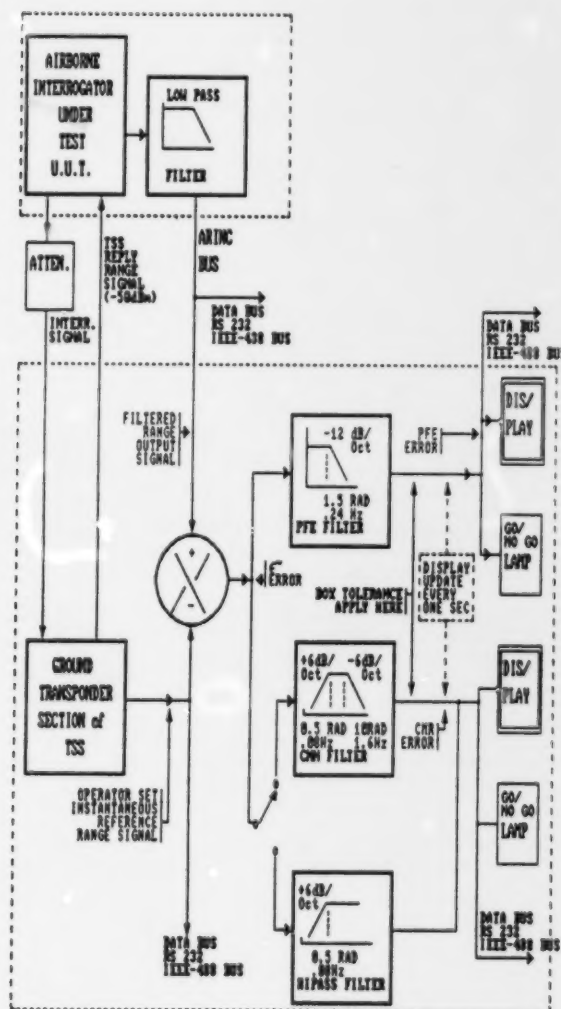


Figure 3. The Test Set Simulator as an Airborne Interrogator Tester

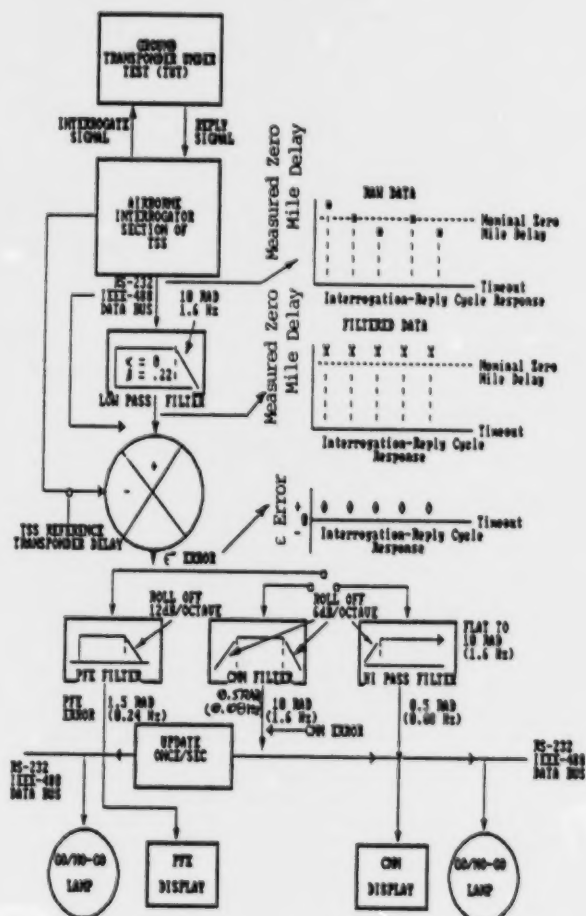


Figure 4. The Test Set Simulator as a Ground Transponder Tester

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

**A DME TRANSPONDER SYSTEM APPROACH FOR
CONVERTIBILITY, MONITORING AND DIAGNOSTICS**

Author: T.A. Hicks
Leigh Instruments Limited
3001 Solandt Road
Kanata, Ontario, Canada
K2K 2N8

Abstract

Both DME/N and DME/P transponders are now current equipment to meet aviation navigational operational use. The transition from DME/N to DME/P design has not been regarded by the industry as a trivial task. This paper shows that by identifying common system requirements, the basis for a high degree of commonality is established for new equipment, and converting between /N and /P systems is relatively easily implemented.

System Performance Requirements

The inter-operability requirement that DME/P transponders must respond to DME/N interrogators seems in effect that the functional capability of both systems is built into a DME/P. In simple terms the primary differences, from a /N perspective, are limitations of power and cost of implementation of the precision DME/P transponder.

The primary interest driving design requirements for DME/P is to improve overall system accuracy to meet the needs of instrumented approach and landing. To accomplish this, minimal timing tolerances are demanded for circuit performance. Critical performance demands for accuracy at landing and takeoff, require solutions for errors introduced by pulse detection timing variance, and also need increased positional update rates. These two concerns have been solved by utilizing a different pulse shape and the separation of interrogation into initial and final approach mode operation.

The major points of performance effecting system design, for a DME/P in comparison to a DME/N transponder are as follows:

- * DME/P must provide service to DME/N interrogators.
- * DME/P accuracy is higher than that of DME/N representing the equivalent, (depending on mode), of 30 ns timing accuracy compared to 300 ns enroute, or 150 ns landing aids, DME/N systems.
- * DME/P uses a faster rise time pulse form, cos squared/cos, instead of gaussian (or very approximately cos/cos) for DME/N. To take maximum advantage of the faster rise time characteristics, different detection thresholds which supplement the half amplitude detection techniques of DME/N are employed. Techniques utilizing 5 to 30 percent rise time linearity and threshold levels combined with delay, attenuate and compare circuits establish a virtual origin time reference for time of arrival.
- * Transmitted power levels to meet absolute power spectrum limits are lower with a /P pulse form. Typically DME/N operates at about 1 Kw (6.3 Kw ERP) for enroute systems, (and at 100 w for local area coverage), and DME/P operates at about 150 w (600 w ERP), providing compatible coverage with operational utilization to 300 Km and 100 Km.
- * Range of operation is limited corresponding to reduction in effected radiated power.
- * To improve final approach accuracy two modes of operation are implemented, initial approach and final approach, interrogation rate is increased for final approach mode. As an expansion of this requirement, priority conditions are enacted for final approach response, in the keying of transponder identifier.

* Characteristics of the receiver (IF), are improved to reduce time delay errors.

* To utilize the /P waveform effectively receiver performance must be enhanced, and in simple terms to achieve time domain improvement, both bandwidth and delay characteristics are to be accommodated. The solutions for an improved receiver with dual receiver paths, impact the IF amplifier which requires both broad and narrow band performance characteristics to meet delay and time reference accuracies and provide channel selectivity.

* Coverage at runway approach levels raises issues of antenna elevation pattern which requires a DME/P antenna solution in critical situations.

* Category 1, 2, and 3 operations impose different performance conditions each category raising levels of reliability and operational confidence as they apply to landing aids. Software simplicity to ensure verification of reliability is currently a major concern for equipment designers.

Functional Commonality

To address these issues in system design and develop a baseline configuration for general DME requirements there must be an appreciation of signal flow, interaction and practical packaging. The functional block diagram for a DME/P transponder shown in Figure 1 shows principle elements of a transponder. Unfilled blocks identify common DME functionality, grey fill, those common functions which are used by both transponder types but are enhanced for DME/P application, and dark fill areas which require specific solutions for DME/P.

From this diagram it is clear that a high proportion of commonality is achievable if it is accepted that enhancements i.e. improve timing performance, for common circuits is applicable to DME/N. This argument can be further advanced if it is seen that by adopting digital timing techniques equipment stability is improved and cost differential is negligible. A detailed analysis of design shows that marginally more circuit area is required for these digital solutions if special LSI components are not implemented. Major differences of function are shown for IF, detector, decoding, pulse modulation, power level and antenna. Since this discussion paper is directed at transponder commonality the antenna impact will only be considered in so far that it provides approximately 2 dB difference in gain which is easily compensated for in both receiver and transmitter signal strength.

Modularity Configuration

A high degree of modularity will allow flexibility to assemble a suite of parts for the transponder function required and will significantly influence the complexity of implementation. Using an approach which takes the basic system functional blocks as the basis for modularity we can match the system requirements for DME/N and DME/P. Exploitation of modularity concepts are the prime thrust for this paper and the process extends from the system requirements into the packaging of the complete system. The requirements of mechanical packaging tend to be neglected by the system designer but this can be the key issue for successful

organization for convertibility implementation.

Figure 2 shows the primary circuit requirements consolidated into logical modular blocks, again identifying those different between /N and /P as dark fill. Logically related functionality and circuit dynamics should be recognized as important in establishing this arrangement.

The diagram shows the implementation to meet the functional differences have been dealt with in different ways.

- * The IF provides dual receiver paths to implement the DME/P solution.
- * To maximize timing accuracies and accommodate high clock rates, the detector and decoder modules group together half amplitude threshold level detection, delay and compare, time of arrival and critical transponder delay timing.
- * Pulse modulation for both transmitter modulator and internal signal generator are located together and both DME/N and DME/P pulse forms can be extracted relatively simply from ROM stored profiles.
- * The transmitter power amplifier can be provided by combining the output from solid state modules. Internal losses in a transponder of 3 dB requires in excess of 2 Kw generated to provide a 1 Kw transponder output power level. Combining three 800 w power module outputs provides 2.4 Kw generated for a 1 Kw DME/N antenna output. For a dualized DME/N configuration the parallel amplifiers are able to maintain equipment availability, and the loss of one amplifier results in a loss of just under 3 dB, (about the alarm point). Using the same 800 w module underdriven to provide 400 w for 150 w DME/P transponder output, no external combiner is required, but to provide backup capability a switch selection to the main unit or the standby is shown.
- * Two control loops are shown which are critical for DME/P but also provide benefits to DME/N. The control of pulse form to maintain linearity, and as a consequence spectrum, is shown as the major loop. The control of delay time requires loop feedback to maintain DME/P accuracies.
- * The diagram expands monitoring elements and introduces an equipment interface.

Performance Monitoring and Diagnostics

Analog and digital monitoring modules are shown communicating with a processor to provide executive monitoring of the transponder. Separated from this group is an interface processor and interface buffer. This is contrived to address the two major concerns of validation of equipment performance and diagnostics. There are different perspectives on the requirements for equipment availability, control flexibility, and the verification of processor software, is particularly vital as it applies to landing aids equipment. The principles of modularity again provide advantageous system benefits. The analog and digital monitors provide determinant hardware circuits for critical parameter measurement. Responsibility for prime transponder monitoring evaluation, and transfer control to a standby set, is embedded in the executive processing. The equipment interface is provided independent of this and is dedicated to diagnostics and interface communications. The communications links are also separated, into a parallel bus used for internal data transfer and control and a serial bus link to the interface unit and remote terminal. This configuration provides a good compromise solution to meet several concerns and provides the flexibility of terminal control of system and diagnostics, separates executive and diagnostics software, but maintains a

reasonable degree of verifiable confidence by providing critical monitoring in hardware.

The consolidation of the functional system requirements into modules as shown, correlating closely with signal flow, is a distinct advantage in system diagnostics. The diagrammatic blocks represent a real modular breakdown where for a complete /N set there are 24 modular blocks and for the /P system 22 modules. Of these /P modules, 5 are sufficiently different to be specific to a /P set.

Transponder Packaging

Finally, in the examination of system configuration, by carrying the modularity concepts further, distinct advantages can be gained by structuring the equipment as a complete elemental structure. Traditional methods of compartmentalization into receiver, transmitter and monitor packages increase interpackage wiring and diagnostic complexity. Using a rack slot arrangement for all modules the configuration of a single or dual (main and standby) transponder can be assembled relatively easily with minimal line length interconnects.

A system using this principle has been developed, which consists of six rows of modules organized into two primary groups, for dual DME/N availability in a standard 19 inch rack. The general arrangement shown in Figure 3 is for a dual DME/N configuration where the standby side is shown in grey fill and changed modules between /N and /P for the main system are shown in dark fill. The upper three rows of modules are organized for the microwave frequency modules and the lower three rows for the IF video and digital signal modules. These rows of individual modules provide easy functional compilation and because of their functional modularity, both simplify diagnostics and allow insertion testing at cardinal transponder system points. This enhances both the remote maintenance monitoring and technician maintenance interfaces for the system.

Acknowledgment

This paper has been developed from DME development program work performed at LEIGH INSTRUMENTS, Kanata, Ontario, Canada, where the author is the Program Manager for DME Development Projects, and responsible for the company funded development of DME/N, DME/P and convertible DME/N to DME/P transponders.

References

Aeronautical Telecommunication ANNEX 10
Vol 1 to the Convention on International Civil Aviation (ICAO)

Robert J. Kelley and Danny R. Cusick "Distance Measuring Equipment and its Evolving Role in Aviation"

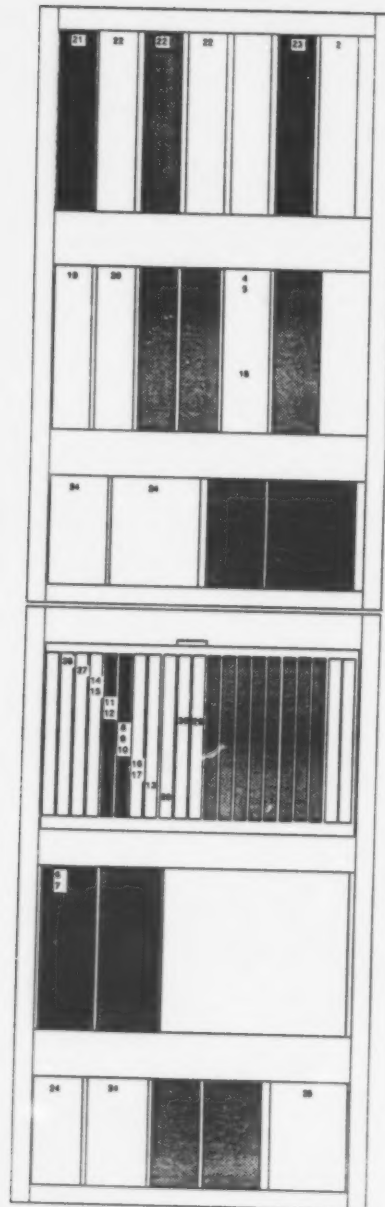


FIGURE 3 DUAL TRANSPONDER PACKAGING

TESTING OF CMA-2000 MICROWAVE LANDING SYSTEM (MLS) AIRBORNE RECEIVER

L. Labrèche & A.J. Murfin
Canadian Marconi Company, Avionics Division
2442 Trenton Ave., Montreal, Quebec
Canada H3P 1Y9

Abstract -

MLS is a precision approach and landing guidance system which provides position information and various air-to-ground data. The position information is provided in a wide coverage sector and is determined by an azimuth angle measurement, an elevation angle measurement and a range measurement (DME/P).

A Technical Standard Order (TSO) has been issued by the U.S. Federal Aviation Administration (FAA) and prescribes the minimum performance standards that the MLS receiver must meet. This TSO defines detailed test procedures for use in determining the required performance under standard and stressed conditions. It also imposes disciplines on software development/testing process.

The testing performed on the CMA-2000 and methods used to validate the MLS receiver are the subject of this presentation.

System Overview

The MLS signal is provided as angle and data functions transmitted in a time division multiplex format on a single radio frequency channel. 200 different MLS channels are defined and are spaced apart by 300 KHz in the range from 5031.0 to 5090.7 MHz. Each MLS function is repeated at a predefined rate including a jitter to provide protection from synchronous interference. Scan functions provide Azimuth and Elevation position information. Preamble codes and data transfers from the ground station are transmitted by Differential Phase Shift Keying (DPSK) of the radio frequency carrier.

Range input to the MLS receiver is provided by other equipment and include DME and TACAN. The receiver may accommodate up to three antennas to enable full coverage during aircraft maneuver. Figure 1 illustrates a perspective view of signal coverage.

Supported MLS ground station site configurations include split-site and other general configuration (e.g. Collocated Azimuth/DME/Elevation transmitters).

Outputs of the MLS receiver to the aircraft systems include Azimuth and Elevation guidance signals and various status annunciators and data. ARINC-429 and MIL-STD-1553 busses provide status and data to other digital avionics.

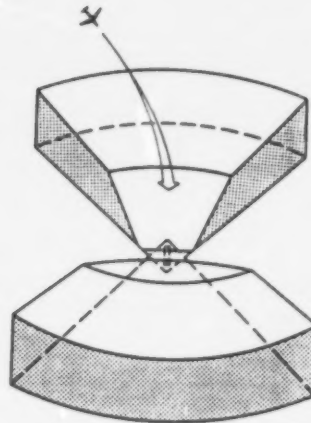


Figure 1: Perspective view of signal coverage

Testing of the CMA-2000 MLS Receiver

Testing of the MLS receiver is mainly governed by Technical Standard Order TSO-C104, which establishes that the receiver performance must be compliant with Section 2 of "RTCA/DO-177 Minimum Operation Performance Standards (MOPS) for MLS Airborne Receiving Equipment" and RTCA/DO-178, "Software Considerations in Airborne Systems and Equipment Certification".

In addition, the Equipment Specification of the CMA-2000 Receiver establishes additional performance characteristics that are not covered by the TSO requirements.

RTCA/DO-177 Tests

MLS guidance accuracy is specified in terms of Path Following Error (PFE) and Control Motion Noise (CMN). These parameters are intended to describe the interaction of the guidance signal with the aircraft in terms which can be directly related to aircraft guidance errors and the flight control system design.

Figure 2 illustrates the components of the Error Spectrum applicable to PFE and CMN for Azimuth and Elevation angles.

The RTCA/DO-177 tests form the basis of equipment performance verification for TSO authorization. Included are performance verification under standard and environ-

mental conditions.

Table 1 enumerates applicable standard condition tests and Table 2 gives a summary of the specified environmental condition tests.

In order to reduce test time, a computer automated system has been developed, which automatically executes the RTCA/DO-177 test cases and verifies all data, status and PFE/CMN output of the equipment. This approach has permitted to perform rapidly a thorough functional verification of the equipment following any software update.

MLS Criticality and Impact on Testing

U.S. Federal Aviation Regulations impose rules on design and testing of critical functions. The probability of occurrence of any critical failure condition in equipment and systems must be less than 10^{-9} per flight hour. The impact of this rule on software translates into extensive verification and traceability effort.

- Extensive verification/testing at each level of the development Life-Cycle (Analytical and test activities).
- Extensive Configuration and Traceability requirements.

Software Verification

Software verification is mainly guided by RTCA/DO-178A. The following testing and verification activities are performed.

- Design Level - Verification of Design against software requirements
 - Compatibility of hardware/software interfaces
 - Completeness and adequacy of system and software requirements
 - System requirements to software requirements traceability
- Module Implementation Level
 - Verification of Module Implementation with standards
- Software Testing
 - Black Box /White Box Testing
 - Testing Methods
 - Modules Testing
 - Module Integration Testing
 - Hardware/Software Integration Testing
 - Requirements Coverage Analysis
- Verification Assurance Matrix

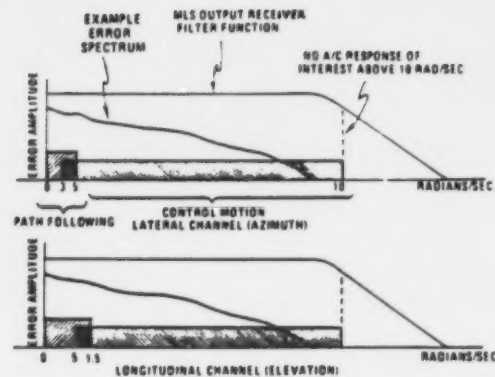


Figure 2: PFE and CMN Components of the Error Spectrum

Table 1: Equipment performance - Standard condition tests

Accuracy without multipath
Warnings
Accuracy in the presence of multipath
Resolution
Output Response Characteristics
Coverage
Clearance
Station Identification
Additional Data
Antenna Performance
Angle Selection
RF Channel Selection

Table 2: Equipment performance - Environmental Conditions

Temperature and altitude
Temperature variation
Humidity
Shock
Vibration
Magnetic Effect
Power Input
Voltage Spikes
Electro-Magnetic Compatibility

Computed Centerline Approach Guidance Testing

The CMA-2000 includes a computed centerline guidance option for which an iterative solution has been implemented.

MLS sensors for Azimuth θ , Elevation ϕ , and Range ρ have inherent errors. In accordance with the accuracy requirements of RTCA/DO-177 and range equipment performance assumptions¹, these errors have been assumed to be:

$$\begin{aligned}ABS(\theta_{error}) &\leq 0.017^\circ \\ABS(\phi_{error}) &\leq 0.017^\circ \\ABS(\rho_{error}) &\leq 0.01 \times \rho\end{aligned}$$

The testing of the algorithm has been performed using a statistical approach based on actually observed outputs in the presence of known errors produced by the sensors (θ_{error} , ϕ_{error} and ρ_{error}).

The solver accuracy goal was established in terms of percentage of random cases for which the algorithm computed the aircraft position within predefined limits and have been established as follows:

For 95% of random cases, the computed approach algorithm shall compute the aircraft position within:

- 1/4 of 1% of the measured range
- 2 feet laterally or 1/4 of θ_{error}
- 2 feet vertically or 1/4 of ϕ_{error}

A PASCAL simulator was implemented on a VAX computer. The simulator was then ported to HP PASCAL for 8086. Over 20000 test cases were run with the simulator driving the actual receiver code.

This test method confirmed the adequacy of the algorithm.

¹ At the time the simulation was performed, the assumption of an error of 1% for range equipment accuracy was considered to be a best case, given the target application, for ranges up to 30 nm. This accuracy goal is however more stringent than the actual accuracy given by the range equipment as the aircraft approaches the runway.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

MLS TESTING USING BUILT-IN TEST

Robert Janes, David Babineau
Micronav Limited
Sydney, Nova Scotia, Canada

Abstract

As Microwave Landing Systems are specified to have extensive Built-In Test (BITE) capability, this aspect of the system's capabilities can be exploited to assist in testing at the factory as well as in the field. This capability does not end at simple functions such as power supply levels and error logging. A particular example of this capability are Micronav MLS antenna systems which include an offline monitor module intended to find failed phase shifters in the array antenna. This module has proven useful in factory testing of the antenna system as a whole as well as phase shifter measurements. A particular example of where this capability was used to advantage is presented.

Introduction

As modern electronic systems, including Microwave Landing Systems, are specified to provide fault isolation to individual modules, Built In Test Equipment (BITE) is a standard part of the system. Use can be made of these capabilities at the factory test level to identify modules that fail during testing and to make measurements that may not be easily accessible as the build of the equipment proceeds.

A case in point in Microwave Landing Systems is the testing of elements of the fully filled phased array antenna. These antenna systems are designed to keep on operating within tolerance despite failures of single elements, so it is difficult to tell if components of the system have failed using the normal parametric maintenance monitors. For this reason, systems using fully filled phased array antennas must have a device capable of detecting failed phase shifters. These are the key elements from a reliability standpoint as they are the only active elements in the phased array antenna. The ability of the phase shifter tests to measure antenna parameters as the sum of the individual parts of the antenna has been exploited to find assembly problems that were not found in the testing of the antenna components.

Phased Array Antenna Testing

The phased array antenna in a Microwave Landing System is shown in Figure 1. It is composed of four major elements:

- 1) Power Divider - This device takes a single RF feed and divides the power between the radiating elements in the appropriate power distribution.
- 2) Phase Shifter - the RF power for an antenna element passes through the phase shifter and its phase is adjusted so that the total of the power from all the antenna elements will form a beam pointing in the desired direction. An array is fully filled if there is a phase shifter for each antenna element.

- 3) Array Antenna Elements - The array antenna is made up of the individual radiating elements of the array. In an Azimuth array for a Microwave Landing system, they are typically each a fixed array of slots cut in waveguide or microstrip patches. Elevation arrays are typically horns or dipoles.

- 4) Manifold Monitor - A portion of the signal in each of the radiating elements is collected so that it sums at a particular angle for monitoring.

The element of the Micronav system that checks the phase shifters is the offline monitor. It can measure the phase and amplitude of each of the phase shifters and flag a maintenance alarm if any are out of specification. The measurement of this data from the manifold monitor also includes the other elements of the antenna. The summation of the phase and amplitude data for the phase shifters produces the phase and amplitude distribution for the whole array. This is a parameter that is difficult to measure with the antenna assembled. It would otherwise have to be done by removing some elements or modules making measurements on the partial assembly and then re-assembling.

Antenna Test Example

The value of the offline monitor as a factory test tool was illustrated by one example where an antenna assembly problem was uncovered on the factory floor. The elevation antenna has an ideal amplitude and phase distribution shaped by the corporate feed as shown in Figure 2a. The phase distribution, when measured at one end of the manifold monitor, has an even phase gradient imposed on it as the signal sampled from each element has a different length of waveguide to travel.

Although the distribution is shaped by the corporate feed, the measured distribution is modified by tolerances on the other antenna transmitter and monitor modules. Even with module testing, faults can occur during assembly which corrupt the overall system performance.

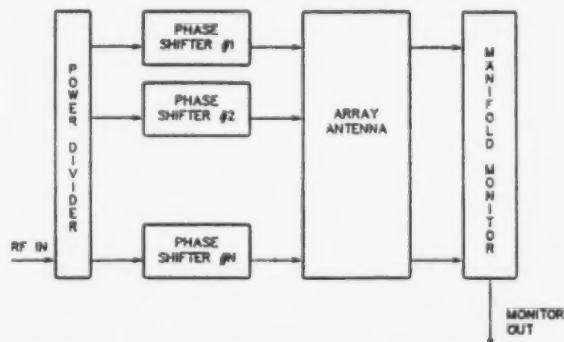
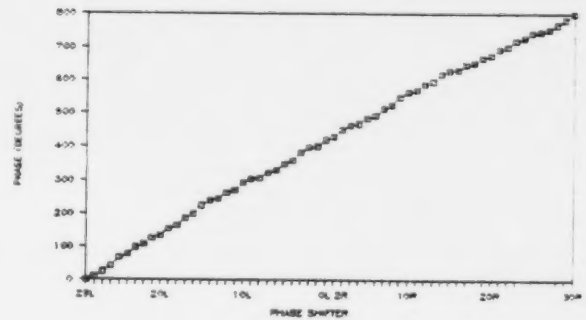
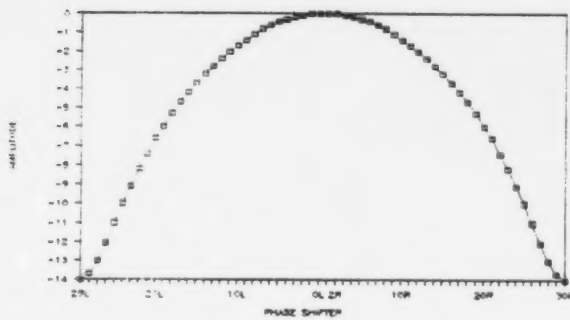
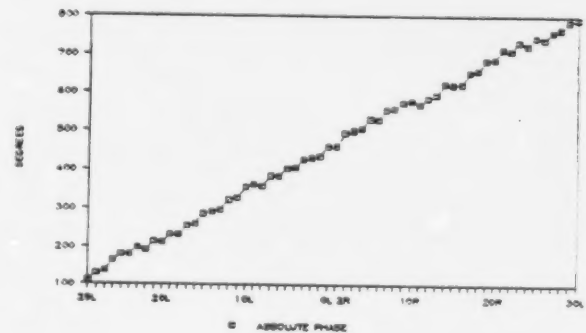
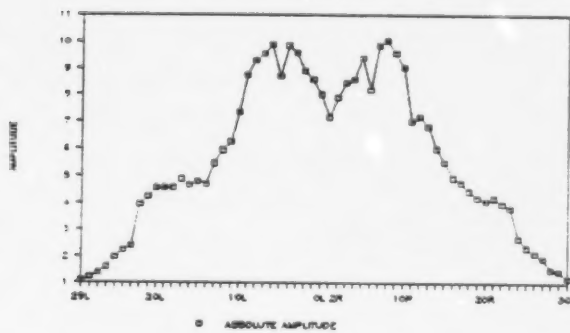


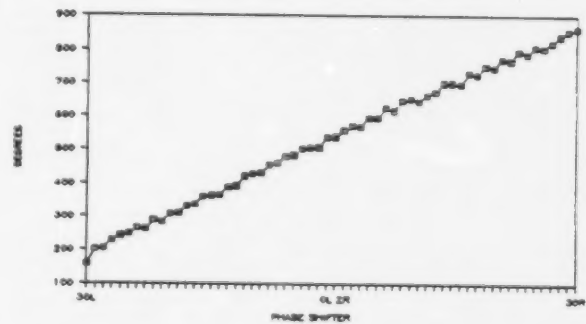
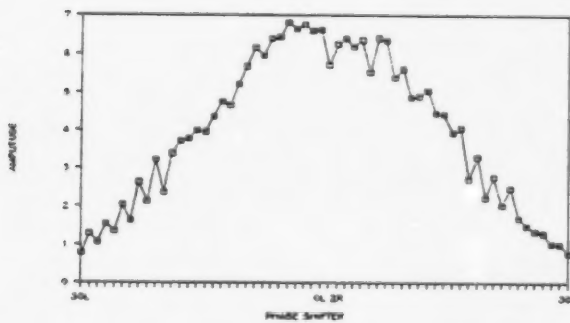
Figure 1 - Phased Array Antenna Block Diagram



a. Ideal Distribution



b. Measured - Poor Load



c. Measured - Improved Load

Figure 2 - Elevation Corporate Feed Test

After final assembly of the elevation antenna, the offline monitor was used to automatically assess the health of the system. The measured amplitude distribution was severely distorted as shown in Figure 2b. It was obvious from the symmetry of the pattern and the breadth of the distortion across the array that the fault was not isolated to certain antenna elements and was affecting the whole array. One parameter which cannot be evaluated by module test and which affects the whole array, is the coupling from each radiating element into the manifold monitor. The manifold monitor is designed to detect a travelling wave at one end and to absorb it at the other. It was postulated, therefore, that if the load was poorly matched, an unwanted standing wave would be set up inside the manifold monitor and distort the amplitude measured from each radiating element. The measurement shown in Figure 2c confirmed this after the match on the load was improved.

The phase distribution measured was not significantly affected by the poor match on the load as shown in the data of Figure 2. Although this fault would not have affected the integrity of the transmitted signal, it would have affected the integrity of the monitor system. This would not have been detected until later stages of testing and would not have been easily found without the offline monitor.

Summary

Systems that have been designed for extensive built-in test have intrinsic capability to provide data for more advanced testing than simply identifying faulty modules. The example of using the offline monitor, normally used for isolating failed phase shifters to do antenna measurements demonstrates the concept. In this case, these problems would have been found at a test range but at a much higher cost to correct. In addition, a computer has replaced expensive, awkward RF equipment and significantly reduced fault detection and isolation time during system testing.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

A STRATEGY FOR EXPLOITING THE FULL POTENTIAL
OF

MLS BASED TERMINAL PROCEDURES
IN CANADA

K. Rami Singh, Eng.

Aviation Planning Services Ltd.
800 Blvd. Rene Levesque, Suite 1420
Montreal, Quebec
H3B 1X9

ABSTRACT

This paper describes the regulatory research conducted in order to define airspace requirements for MLS based terminal procedures. In particular, the extent of the research and the time frame for the translation of findings into regulation/standards is addressed. Constraints which have prevented fully exploiting the potential of MLS based terminal procedures by commercial air carriers are discussed, followed by a research strategy which could result in a more timely realisation of the MLS promise in Canada.

INTRODUCTION

In addition to the research leading to the selection by ICAO of one standard MLS system, the Time Reference Scanning Beam (TRSB) in 1978, a great deal of regulatory thinking and research has been undertaken by some governments and private agencies to further the concept and implementation of MLS based terminal procedures (TERPS) and achieve the promised benefits.

The promises of the MLS in comparison to the ILS have included:

- (i) A lower cost installation.
- (ii) Reduced requirement for airport land.
- (iii) Availability of more channels.
- (iv) More precise guidance in terminal area through to touchdown, flare, rollout and turnoff (leading to possibly lower landing minima).
- (v) Relatively error and interference free signals in space in areas of terrain, structural reflections or FM radiation.
- (vi) A wider coverage volume in space.
- (vii) Considerably more complex approaches such as segmented and curved path in the horizontal plane and variable glide slopes in the vertical plane.
- (viii) Multiple approach paths with variable final approach segments for different speed aircraft to improve flow capacity at airports.

There does not appear to be any disagreement on the potential of the MLS to deliver on most of the above promises, given an adequate time frame and funding for research. The only exception may be the claim that MLS will provide lower landing minima than an ILS would. There is, however, concern that MLS is not the only means of achieving some of the aforementioned capabilities. This concern is motivated by industry experience with relatively complex procedures using conventional technology/navigation and the absence of research and/or literature on applications with the current technology, equipment and procedures.

It is not the intention here to prove or disprove the potential of the MLS to deliver any specific capabilities or to debate the relative merits of ILS versus MLS for the global community. Rather, the objective here is to briefly explore some of the research and the incremental capabilities this research is intended to bring to the operations of commercial air carriers and/or facilitate more through-put in the Air Traffic Control System. Also included is a discussion of the concerns voiced by the user community, which include in particular the pilot group, a review of some unusual but current operations for comparison with the MLS promises, and a suggested strategy for exploiting current technology, operational philosophy and experience to achieve a faster realisation of the MLS promise for Canadian applications.

Therefore, the focus of this paper is primarily on operational issues. The discussion of the research conducted to date is divided into a pre-MLS selection period and a post-MLS selection period. This is followed by the review of industry concerns, Canada specific applications and issues, and a proposed research strategy for Canada.

PRE-MLS SELECTION RESEARCH

The ILS has been the dominant precision aid for some time. A major effort to define the approach and missed approach airspace requirements for the ILS involved the ICAO collision risk model (CRM). The work leading to this model is well documented in ICAO references.¹ For the ILS CRM, radar position data for aircraft on ILS approaches (final phase) was collected and used to define the clearance planes in the vicinity of the runway. For the other types of approaches such as non-precision NDB, VOR, DME ARC, Localizer only, offset localizer, and terminal area transitions, airspace requirements have been defined in other ICAO documents (PANS-OPS), although literature on the research leading to the definition of these requirements has been sparse.

With respect to the MLS, from among the systems proposed by Australia, Britain, France, West Germany and the USA, the TRSB system of the US was selected by ICAO as the standard in MLS after substantial demonstrations by NASA in 1976, with their Terminal Configured Vehicle (TCV), a modified Boeing 737.² The demonstrations were intended to show only capability and were not used specifically to collect data from which standards could be derived. They included "area navigation (RNAV) flights into the TRSB coverage area, transition to TRSB guidance, curved descending approaches to a 3 mile final approach leg, and an automatic landing with a rollout under TRSB guidance." Two types of approaches, an "S" turn and a 130° turn to final were also demonstrated.

POST-MLS SELECTION

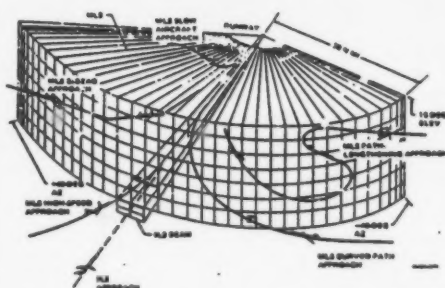
The long term interest of most of the MLS research has been the capability of achieving automated landings after arbitrarily complex or different approaches to the runway. The expectations of the types of approaches to be offered by the MLS are displayed graphically in figure 1, taken from the work of J. B. Feather.³

While some data has been collected on the human factors such as pilot perception and performance during complex approaches under MLS,⁴ and the workload analysis under exclusively MLS scenarios,⁵ the bulk of the research to date has been accumulated on procedures which essentially replicate ILS based configurations and would allow Category III limits (precision guidance down to touchdown and rollout under zero visibility conditions). These have been straight in precision guided approaches and straight out un-guided missed approaches.

Variations of glide slope have been extensively tested in a number of aircraft types, representative of the spectrum of general and commercial aviation aircraft. These aircraft types have included Boeing 727, Boeing 737, Lockheed C-141A, Sabre-80, Convair 580, Beechcraft 200, and a Cessna 172.⁶ However, the collision risk model, which reflects the most precision in airspace requirements development, has only been defined for the 3° glide slope.⁷

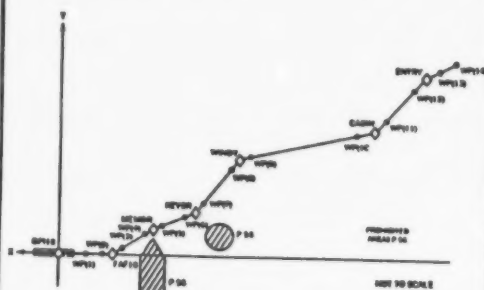
Some testing has been done on "complex" approaches using MLS waypoints. A graphical representation of one such approach is shown in figure 2, which has been reproduced from another paper by J.B. Feather.⁸ It portrays the IFR version (using MLS Area Nav) of the Washington National River Approach. However, this testing was done in an MD-80 simulation program.

**FIGURE 1
POSSIBLE APPROACHES IN THE
MLS COVERAGE VOLUME**



SOURCE: J.B. FEATHER

**FIGURE 2
WAYPOINTS FOR WASHINGTON
NATIONAL RIVER APPROACH**



SOURCE: J.B. FEATHER

The National Research Laboratories of the Netherlands has also conducted some research, using an actual aircraft. B-747 simulation tests were conducted utilizing 2.0 to 4.5 nm finals from turn. Other aspects studied included glide path interception and radius of turn to final. This data has been made available to the FAA, who plan to integrate it into its standards development.

In addition to straight in approaches at various glide path angles, the FAA has done some data collection on segmented path (SP) and curved path (CP) MLS based approaches. The resultant MLS criteria which have emerged (FAA Order 8260.36⁹) allow glide slopes up to 6.4° according to the aircraft category. Aircraft of category C, which include B-C-9 size aircraft, are limited to maximum glide paths of 3.6°. Category D & E aircraft are limited to 3.1°. This includes B-747, L-1011 size aircraft. In terms of the curved and segmented path criteria, the interesting clause is that these can only be established in areas of terminal radar coverage, which implies that Air Traffic Control would either automatically or manually monitor the progress of aircraft and would provide the necessary back-up, should there be a failure of a component of the MLS system.

Although the FAA (and NASA) have not been able to conduct research on all of the "creative" applications of MLS, the MLS Implementation Policy, documented in the White Papers of the FAA¹⁰, offers the following statement:

"Continue to identify and encourage individual opportunities to install MLS ground systems at locations that can deliver significant benefits and/or offer early and unique opportunities to take advantage of MLS capabilities."

Transport Canada (TC) has also been active in exploiting unique applications of MLS and has, for example, expended considerable resources in the development of the MLS based, segmented approach procedures at Pemberton Airport, B.C. Other airports have been studied either with Government funding or direct TC involvement. These include Port Alberni, Revelstoke, Toronto Island, and Jasper-Hinton, where an MLS procedure has been in existence for three years.

The regulatory tack taken by the FAA with respect to the "unique opportunities" is that of certification on an exception basis, as appears to be the case at Aspen, Colorado. The factors taken into account in exception certification include, in addition to the location specific guidance signal, aircraft performance, air carrier and crew training. The case of approaches where the final segment is similar to the conventional ILS final segment, some MLS applications have been in effect for some time such, for example Valdez, Alaska. Where the procedures are very different, and involve, for example, mountain valley following with segmented paths, progress has been slow.

To respond to the issue of alternative precision guidance systems exclusive of ILS and MLS, the FAA has addressed the proposed alternative, the Global Positioning System (GPS). Considerable theoretical analysis has been done and presented among the white papers published by the FAA. In a recent interview with FAA officials¹¹, it was learned that further tests are planned, using a Saber-80 equipped with differential GPS. These tests will involve initially non-precision approach tracking accuracies, followed by precision approach tracking accuracies. In any case, the analysis documented in the white papers suggests that the GPS cannot offer the required (ICAD) precision for Category II and III approaches, and can offer Category I precision only under specific conditions of service by the US. In addition, a case is made for the risks inherent in competing with the interests of US National Security, which would determine the GPS signal's degradation to be intentionally introduced into the system.

Other areas of research have included transition to MLS from other navigation sources¹², and marriage of MLS and area navigation (such as INS, LORAN-C). However, all of the documented research appears to have been based on all systems functioning within specifications with no failures of electronic, powerplant or human components. However, the FAA plans for a wide body simulation¹³ to specify that:

"The only malfunction required to be simulated is the loss of MLS guidance during departure or an approach. When failure occurs the area navigation system guidance should automatically revert to VOR/DME and air data (barometric altitude) to compute position estimates. The capability to model other insidious failures, such as incorrect waypoint storage in the data base, and incorrect aircraft position determination is required to assess pilot sensitivity and performance in the presence of such failures."

This aspect of the research is also of concern to the industry in general and the pilots in particular¹⁴. The absence of some types of failures such as powerplant outages on aircraft, which constitute a major portion of the training of career pilots, is an example of one such concern.

INDUSTRY CONCERNS AND PROPOSED RESEARCH

The reservations about the FAA plan to replace ILS with MLS have been documented in reports by the Industry MLS Evaluation Task Force¹⁵. It is believed by the US airlines and others in the aviation community that

"Recent technical improvements to ILS,...., sophisticated airborne systems installed on new aircraft types will enable airline and other similarly equipped aircraft to realize whatever capacity enhancement benefits that may derive from the ability to fly curved or segmented approach paths without reliance upon MLS as the signal source for the curved approach path...".¹⁶

A joint FAA/Industry MLS/ILS Assessment Project has been set up to conduct nine specific projects¹⁷. This will involve the following sub-projects:

- (i) Analysis of available ILS frequencies for expansion of precision approaches in the US.
- (ii) Evaluation of wide body curved or segmented approaches and other advanced technologies for precision approach.

This project in particular will explore the MLS RNAV and FMS/ILS curved and segmented approach procedures, flight crew and passenger acceptance of these procedures, as well as other new aviation technologies.

- (iii) Evaluation of advanced procedures in multi-airport environments.

This project will attempt to quantify and document the effects of advanced curved and segmented approaches.

- (iv) General Aviation/Commuter Capacity Enhancement

This project is motivated by the observation that under VFR conditions, airport flow capacity is at its optimum, despite the mix of different category aircraft, while, under IFR, this capacity deteriorates. It is intended to quantify the capacity effects derived from the availability of MLS approach facilities at commuter and general aviation runways, and optimised air and ground circulation patterns, which will enable segregation of Category A and B aircraft from heavier aircraft.

- (v) Comparison of MLS to ILS performance

This project will assess the comparative performance of MLS versus ILS in such areas as accuracy, low altitude coverage, reduction of overflight effects, reduced critical areas, immunity to weather effects and flyability.

- (vi) Assessment of MLS avionics installation costs

- (vii) Assessment of reduced MLS minima

It has been asserted that lower landing minima are possible with MLS due to the notable accuracy of the MLS-DME/P. This project will collect and analyze data that will support reduced minima, including development of standards for cockpit procedures, aircraft systems, runway visual environment, and related operations specifications as well as the cost of airborne equipment.

- (viii) Development of DME/P Interrogators

Currently, there are no airborne interrogators that are designed to advantage of the improved accuracy that DME/P will provide. Therefore, the benefits of these units cannot be demonstrated. This project will fund development of such units for demonstration purposes.

- (ix) Category II/III flight demonstration and Procurement

This will involve integration of MLS-DME/P avionics with flight directors and autopilot systems to provide Category II capability in turboprop and separate Category II/III capability in large corporate or air carrier jet.

The schedule of these demonstration projects shows a final completion date of the end of 1991. It is expected that the results from these projects would lead to reasonably rapid resolution of the current concerns within the US and implementation. However, Canadian interests are possibly more modest than those of the US industry.

ASSESSMENT OF PROJECTS FROM A CANADIAN PERSPECTIVE

From among the above list, projects on CP/SP approaches, advanced procedures at multi-airport metropolitan areas, general aviation/commuter capacity enhancement, and assessment of reduced MLS minima are of particular interest from an operational point of view. From a Canadian perspective, it is interesting to note that at airports where capacity problems have become intolerable, for example, at Toronto Lester B. Pearson Airport (LBPA), and Vancouver, proposed solutions have rarely implied the need for a different precision guidance system or complex approaches¹⁸.

Solutions to capacity problems have included construction of more runways and taxiways, extension of hours of operation, change in noise abatement restrictions, air traffic control staffing increases, air traffic flow management compatible with the US, and separation of categories of aircraft, i.e. slow speed aircraft on separate runways or airports, and split runway operation. With the exception of Toronto Island Airport, where constraints on airspace usage under IFR conditions are more due to obstructions than proximity to other airports, multi-airport problems such as the ones in the New York are not of serious concern at the moment. Perhaps, as part of the solution to the saturation problem at LBPA, other airports may be involved. However, these airports are not expected to be in the same proximity as La Guardia is to JFK.

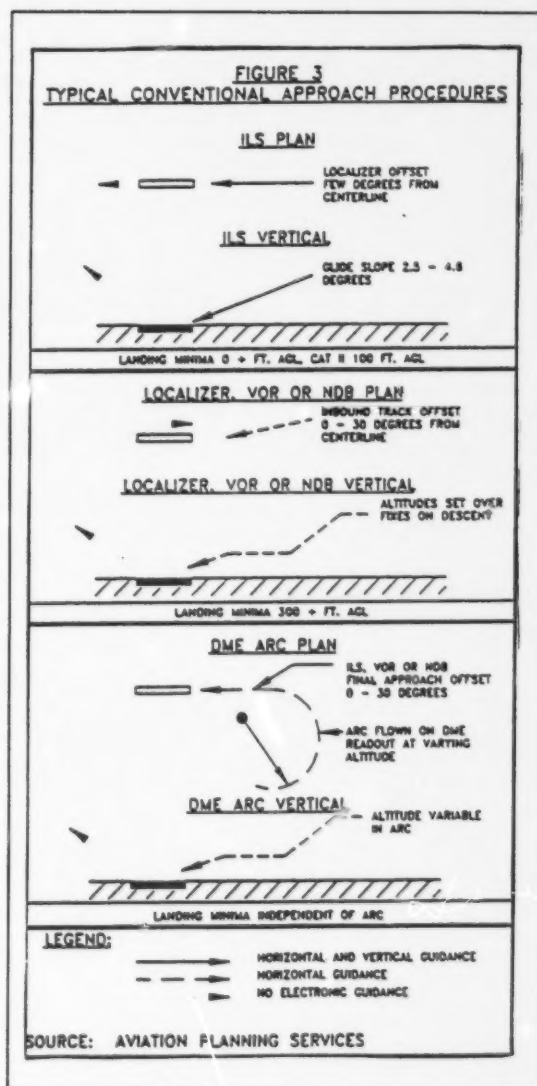
The net implication to Canada is that of all of the potential applications, the ones which could be most beneficial would be the ones which would lower the landing minima at current airports or provide IFR capability at airports which do not currently have such service. Both these capabilities are associated with the ability of the guidance systems to provide reliable and reasonably precise signals for terminal navigation and established procedures which provide adequate terrain or obstruction clearance.

The cumulative experience around the world with conventional nav aids (ILS and the other non-precision nav aids) has led to the development of a wide range of conventional approaches. Figure 3 shows two types of "generic approaches" practised in the world, precision and non-precision, and one option for manoeuvring onto the final approach path, the DME ARC transition.

Among the precision approaches, it should be noted that Toronto, Montreal, Halifax, St. John's, Winnipeg, and Vancouver already have Category II approaches based on ILS. This provides landing minima of 100 feet or approach end Runway Visual Range (RVR) of 1200 feet. Slightly higher minima (200 feet or RVR 2600 feet) are available at almost all Canadian non-mountainous airports where significant air service is maintained. With this type of precision, airport closures are rare due to low visibility or ceiling alone. However, the other major hindrance to operations in Canada does force service disruptions. This is the case of thunderstorms and contaminated runways due to ice, snow or standing water, which when coupled with crosswinds create more havoc with schedules than the absence of precision guidance.

In the case of thunderstorms, no navigational guidance system, regardless of precision, could convince pilots to land or takeoff. With respect to contaminated runways, the manoeuvres executed by a pilot, particularly in crosswinds, where the aircraft may tend to weather cock and appear to drift off the runway, include deliberately firm touchdowns "on the numbers" followed by delicate manipulation of the rudders, ailerons and cycling between forward thrust and reverse thrust as required to maintain the aircraft within the runway confines. The pilots sensitivity to the aircraft response is enhanced when a he/she establishes visual contact with the runway and almost instinctively assess the drift, required bank angle for centerline tracking, and braking action. On conventional approaches, this occurs below 200 feet or so above touchdown height and by the time the threshold is crossed, the aircraft attitude for touchdown is more or less established.

Without a significant visual segment on the approach, pilots would be reluctant to touchdown on a runway they cannot see clearly in these circumstances. Auto-land capability, which is becoming more common in newer generation aircraft, has not been enhanced with the "seat of the pants" control responses described here. Therefore, in these weather conditions, auto-land may not be allowed to conduct landings, despite the precision of the guidance system. Even if Category III landings were possible, at busy



airports, ground control under zero visibility conditions would pose another problem which has not yet been addressed adequately. The conclusion from the above discussion is that at Canada's busiest airports, capacity cannot be significantly improved by lower landing minima from those currently available.

In terms of IFR service to other Canadian airports, among the non-precision approaches, it is common to have landing minima in the vicinity of 300 feet for Localizer and 400 feet for NDB, if terrain or obstructions are not present. Most non-mountainous Canadian airports have at least a non-precision approach of some configuration.

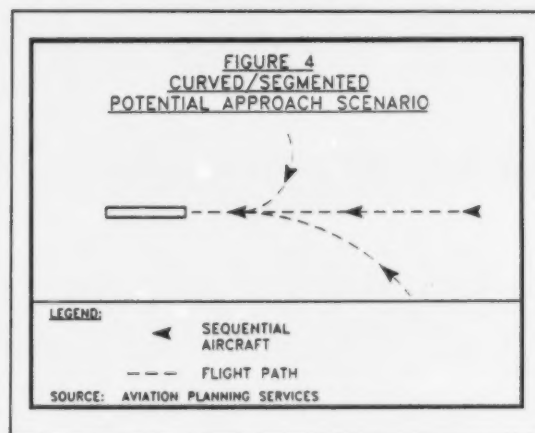
IMPLICATIONS FOR CANADIAN CAPACITY ENHANCEMENTS

The net observation from the above discussion is that in Canada, lower landing minima than the 100 or 200 feet would not necessarily increase flow capacity at the airports where capacity problems have been identified. On the other hand, airports, where landing minima down to at least 300 or 400 feet would allow a significant improvement in airport usability, have been identified. This includes airports such as Toronto Island. Also airports located within valleys, where no instrument procedures currently exist, have

also been identified. Here, any IFR procedure, regardless of landing minima would enhance capacity.

On parallel runway airports such as LBSA and Dorval, increased capacity could be realised with multiple and parallel operations. In this respect, any improved accuracies of MLS over conventional ILS would be beneficial. However, the extent of this improvement is not significant in view of the findings of Dr. Yates¹⁹. Furthermore, the restrictions at LBSA on multiple runway operations are also noise motivated. It is tempting to suggest that segmented or curved approaches could be possible solutions. However, the final capacity constraint on single runway operations is the separation between aircraft of various categories on final approach, the time to clear the runway, and separation between landing/departing aircraft. This is independent of the guidance system.

Another concern with respect to curved and segmented approaches in high density airspace, although not unique to Canada, involves the protected airspace around each aircraft while on a sequenced approach. Figure 4 shows a potential scenario. If maximum throughput is the objective, such a



scenario is possible if the aircraft can be made to fly ground speeds rather than airspeeds. This will ensure the minimum separation is maintained at the touchdown point. However, current operations are conducted strictly to indicated airspeeds, necessitating frequent controller intervention to ensure the required separation. Under the scenario depicted in figure 4, with a modest wind vector field, the task of maintaining constant minimum separation is clearly non trivial.

A further concern during curved and segmented approaches in high density airspace is the impact of electronic or powerplant failures. The FAA order 8260.36 requires that such procedures be permitted only under a radar environment. This provides a backup in case one aircraft loses navigation capability. If the ground based system loses a critical component, the task of delivering aircraft to the runway becomes more complex and risky, despite radar control, given the configuration and state of the terminal airspace when manual control may be assumed. Some industry experts have suggested backup navigation systems such as conventional RNAV or INS/FMS systems be mandatory while conducting such procedures. Others have gone on to suggest that since these conventional systems currently are dual installed on most commercial aircraft, and the accuracy is very high²⁰, why not define such procedures independent of a ground based precision signal.

The procedure shown in figure 2, could, for example be flown entirely by conventional RNAV and VNAV functions to limits from which a visual segment could be flown. In Canada, we do not as yet have procedures comparable to the Canal approach in New York, or the Washington River approach. However, scientific accuracy estimates, not in relative terms (to MLS), but in absolute terms, have yet to be determined. For Canadian purposes, these may be adequate. It may be of interest to note that an IFR procedure dependent strictly on conventional VOR/DME based RNAV has recently been authorised at St. Jean, Que.

In any case, backups of MLS with RNAV or INS, RNAV with RNAV, or some other combination, will help tracking accuracy under navigation system failures. However, if we introduce powerplant or hydraulic system failures on an aircraft during high density CP/SP scenarios, control may involve an entirely random path for a short time. The potential repercussions of such failures clearly need to be defined and studied before any implementation takes place.

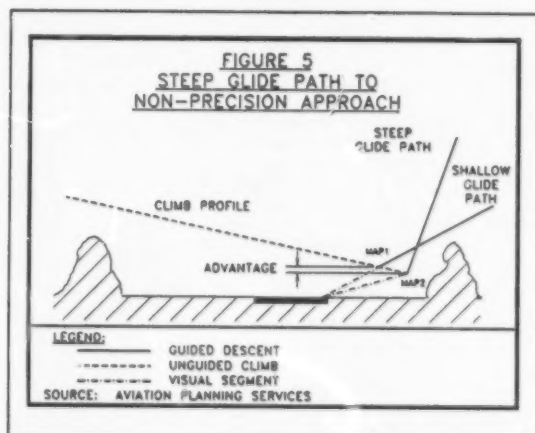
At present, the standards for the approaches have not been based on assumptions of failures and degradation of lateral tracking capability, although take-off and missed approach clearance planes have been based on very conservative engine out performance considerations. Current procedures development has not accounted for multiple aircraft converging on paths whose tangents may intersect in time and space. Therefore, implications for capacity enhancement from SP/CP approaches are not at all clear.

IMPLICATIONS FOR NEW IFR SERVICES IN CANADA

The promise of reliable precision MLS guidance in mountainous or urban airports which could not have conventional navigation aids due to terrain or obstructions, has brought on extensive interest from small communities in the Rockies and carriers serving or wishing to serve these communities with reliable air service.

In most of these community airports, the problem has not only been one of inadequate airspace for the approach, but also the absence of enough vertical and lateral clearance on a potential missed approach, given the limitations of the assumed aircraft performance in current standards. Part of the problem has been resolved by a generation of aircraft which can make steep approaches and relatively steeper climb-outs with one engine failed. However, even this has not resulted in the fulfilment of the MLS promise at some airports where this generation of aircraft cannot provide the load or the range capability.

There are a few options which can be explored scientifically in the interest of achieving the MLS promise in Canada. It has been found in many planning exercises for these airports that a significant degree of IFR service reliability can be achieved by providing instrument procedures which do not necessarily align the aircraft and place it on a landing glide path in relation to the runway but bring it safely to a point from which a visual landing can be made. Traditionally this point has been on the flight path which continues un-interrupted to the runway touchdown. Attempts to define precision approaches using conventional glide paths to touch down have failed in these airports. Figure 5 shows possible reasons for the failure of such attempts.



In figure 5, the constraining flight path is the missed approach path. The conventional shallow (or steep approach) which brings the aircraft to touchdown could descend only to the point of intersection with the missed approach path. This intersection occurs at a higher elevation than one which brings the aircraft to a hypothetical point prior to the runway. The

difference in these points is labelled the "advantage." It can be seen that lower minima can be achieved if this strategy is used. In some cases, the terrain may not allow the glide path to intersect with the runway. In this case, no instrument procedure has been permitted. Airports such as Port Alberni and Revelstoke are examples of this phenomena. Currently, there are no procedures that use this strategy in Canada. Furthermore, there does not appear to be any research on the use of this hypothetical touchdown point before the runway. The most advanced procedure contemplated in Canada, and possibly anywhere, has been developed for Pemberton but is awaiting further testing and weather analysis for licensing. This method of exploiting the MLS promise in Canada merits further investigation.

A related issue has been the ability of conventional aircraft to fly glide slopes steeper than the 3.6°. FAA limits have been referenced earlier in this paper. Effectively, all aircraft approximately equal to or larger than the B-727 could not fly steeper approaches. The "heavy" aircraft could not exceed 3° glide slopes. By contrast, in Canada, aircraft of all sizes, including B-747 and L-1011 have used the Stephenville airport in Newfoundland under instrument conditions with an ILS glide slope of 4.5° and landing minima established at 416 feet above ground level. Another example is the ILS approach at Chamberry, in a mountainous region of France, where a glide slope of 4.6° has been provided for Category A,B, and C aircraft.

Operational experience with approaches at Stephenville airport, with heavy jets, indicates that while the descent rate on the approach is higher than normal, from 420 feet above ground, a relatively normal landing can be made by adjusting the descent rate and touching down beyond the glide path and runway interception point.

Another type of operational experience pilots often have is the need to expedite the descent when caught too high on a final approach. This can occur in aircraft of all categories. The interesting aspect of this experience is the vertical speed which is attained, the proximity to the runway, and the three dimensional navigational guidance. Vertical speeds of approximately three times the normal vertical speed (3° glide slope) have been achieved to as close as four miles from the touchdown point, with only lateral tracking guidance, a DME and vertical deviation indication. On aircraft equipped with VNAV and LNAV, this manoeuvre can be automatically conducted with adequate precision. On other aircraft, precision is lacking but the procedure has been accomplished successfully with predictable regularity.

The observation from the above experiences is that, provided the lower limit of descent is sufficiently high, steep descents can be accomplished by all types of conventional aircraft to limits at which a visual or instrument transition to an acceptable descent rate and landing can be conducted. If we also accept that the previously referenced "hypothetical" touchdown point could lower IFR descent minima, then this opens up further opportunities for the MLS to provide or improve service in restricted areas of Canada, either from terrain, noise abatement, or missed approach obstruction clearance points of view.

In the literature, the fixation has been with achieving potential Category III capability on every installation and the vast intermediate applications, especially in Canada and many other countries with similar topography and traffic densities, have not been addressed. Therefore, waiting for others to be imaginative and do the research could mean significant delays to the realisation of the MLS promise in Canada.

A RESEARCH STRATEGY FOR CANADA

It should be pointed out at the outset that Transport Canada (TC) and the Transportation Development Centre (TDC) has been very active in research involving the unique application of MLS at Pemberton and other studies of economic and technical feasibility elsewhere. In addition, the electronics, avionics, manufacturing, installation, and operations of the ground based systems have been very carefully followed by appropriate TC officials.

The proposal for this research strategy for Canada stems from

- (1) an observation that much of what is currently done has not been adequately studied and its potential explored, and
- (1f) an assumption that "unique" approaches can be built up from modules, which may consist of conventional and new technology components, and which may be studied and tested

in detail individually and then collectively as a total system. This is certainly the case with hardware and software components.

In terms of MLS type applications, the research strategy implies the determination of, among other elements, the following :

- (i) Distribution of tracking errors under failures or reversion to manual for a wide variety of procedures.
- (ii) Functional relationship between lateral airspace requirement and the angular change in track, DME distance, and glide slope angle.
- (iii) Functional relationship between vertical deviations and the glide slope angle, DME distance, and prescribed profiles.
- (iv) Distribution of vertical and horizontal profiles for steep precision guided approaches terminating in non-precision or visual final approach segments.

With these relationships, pieces may be joined to establish acceptable terminal procedures which satisfy some airport service objectives. However, the determination of each of these elements requires further consideration. Current research in the USA and the Netherlands has been limited to a few aircraft types, tests conducted within the total system under actual conditions and some flight simulator studies on even fewer aircraft types. Therefore, contributions of individual components to errors in the system have been difficult or impossible to isolate. The modular approach of procedures development has not been advanced.

Very briefly, the errors of interest have been those due to the MLS (or any navaid system), the installation of the ground systems, the signal in space and RF errors, the avionics receivers and displays, antenna positioning, external factors particular to the installation, and the flight tracking errors. Many of these errors can be reduced by specification and tested independent of the presence of any aircraft. For example, ICAD Annex 10, defines the acceptable tolerances for the MLS hardware. Other components can be flight tested by regulatory agencies. These include signal strength and reliability of signal in space. By the time a commercial carrier comes to use a procedure, the only remaining errors are those introduced by the operating procedures and the weather environment (wind and turbulence). These errors, in particular, are the ones which the present strategy proposes to quantify.

It is theoretically possible to test a wide variety of aircraft types, pilot compositions, failures, wind, turbulence and procedural, and system components in flight simulators. Data on deviations is attainable with little software modifications. Given the distributions of other errors, either by specification, or from tests with flight inspection aircraft, it is theoretically possible to develop convoluted distributions which describe total system errors. For a few actual aircraft flight tests, if these convoluted distributions can be validated, then the results from the wider range of simulations could be considered representative of the population. This approach would provide the extensive base of procedural and failure based data needed to construct arbitrarily "unique" MLS type approaches, which appear to be the ones needed most in Canada.

CONCLUSIONS

While a great deal of research has been conducted, and is planned, to put in place a set of standards which would solve many of the airport capacity, noise abatement, terrain clearance and landing minima problems, from a Canadian point of view, only a few of the capabilities promised by the MLS are needed. These are, however, low on the priorities list of the most significant contributor to the research efforts in the world. Specifically, what are termed as "unique" applications by the FAA are needed for Canadian interests. It is, therefore, necessary to adopt a research strategy that would lead to a faster realization of the MLS promise for Canada.

Much of this research work can be accomplished in flight simulators, in Canada, coupled with theoretical analysis and some flight validation with appropriately equipped Transport Canada aircraft. The results would expedite the development of standards which, in Canada, remain at the mercy of the United States or subject to non-comprehensive development.

REFERENCES

1. Manual on the Use of the Collision Risk Model (CRM) for ILS Operations, Doc 9374-AN/504, First Edition, 1980, "List of References," pages (vii) and (viii).
2. White, William F. "Flight Demonstrations of Curved, Descending Approaches and Automatic Landings using the Time Reference Scanning Beam Guidance," NASA Technical Memorandum 78745, May 1978.
3. J. B. Feather, Guidance Studies for Curved, Descending Approaches Using the Microwave Landing System (MLS), NASA Contractor Report 178030, May 1986.
4. Richard DeLoach and Jacob A. Houck, Pilot Evaluation of experimental Flight Trajectories in the Near Terminal Area, AIAA Atmospheric Flight Mechanics Conference, Williamsburg, Virginia, August 18-20, 1984.
5. Leland G. Summers, Crv Procedures for Microwave Landing System Operations, NASA Contractor Report 178389, Langley Research Centre, Hampton, Virginia, November 1987.
6. FAA reports numbers : FAA-AVW-306-30, 31, 32, 33, 34, ..., Aircraft Type Microwave Landing System Flight Test Program Reports, Aviation Standards, National Field Office, Oklahoma City, OK, 1986 and 1987.
7. J. W. Yates and Allan Jones, Statistical Partitioning and Probability Distributions for Microwave Landing Systems, Unpublished report, Standards Development Branch, FAA, Oklahoma City.
8. J. B. Feather, Guidance Law Simulation Studies for Complex Approaches Using the Microwave Landing System (MLS), NASA Contractor Report 178182, November 1986.
9. US Department of Transportation, Federal Aviation Administration, Order 8240.36, 11 July 1988.
10. Associate Administrator for Aviation Standards, Microwave Landing System (MLS) White Papers, U.S. Department of Transportation, Federal Aviation Administration, May 1988.
11. Interview by author of Standards Development Branch, FAA Oklahoma City, 21 July 1989.
12. R. W. Hoeschen, Flight Tests of Three-Dimensional Path-Redefinition Algorithms for Transition from Radio Navigation to Microwave Landing System When Flying an Aircraft on Autopilot, NASA Technical Memorandum 4089, Langley Research Centre, January 1989.
13. Unpublished internal document "Task Description MLS Wide Body Simulation," FAA, 1989.
14. Previously referenced document of the Air Transport Association of America.
15. Joint FAA/Industry MLS/ILS Assessment Project, Air Transport Association (ATA) of America document, 30 March 1989.
16. Attachment 2, of ATA internal memorandum, 30 March 1989.
17. Previous reference.
18. Air Transport Association of Canada (ATAC) Newsletter, April 1989.
19. Referenced previously.
20. Accuracy measured subjective visual evaluation of the position of the aircraft on approach after descending below cloud on a PBN/RNAV/INS non-precision procedures.

NEAR FIELD RECEIVING PROPERTIES OF PRINTED LOOP ANTENNAS

Z.A. Delecki*, M. Barakat*, H. Kroker*, L. Shafai**, M. Pawlak** and L. Soiferman***

- * National Research Council of Canada, Division of Electrical Engineering, 435 Ellice Avenue, Winnipeg, Canada, R3B 1Y6
- ** The University of Manitoba, Department of Electrical Engineering, Winnipeg, Canada, R3T 2N2
- *** Wardrop Engineering Inc., 435 Ellice Avenue, Winnipeg, Canada, R3B 1Y6

Abstract

In this paper, the receiving properties of printed loop and spiral antennas are considered. The theoretical basis is formulated for the solution of this type of antennas. The general theory is formulated and specialized to the rectangular printed spiral. An example of experimental results is presented.

1. Introduction

Knowledge about performance of the printed loop and spiral antennas in the near field is important in many areas of engineering. In many instances, the experimental design leads to large number of applications, [1]. Most of the antennas were designed for the far field performance. The possibility of applying this type of antennas for contactless testing, e.g. signature diagnosis of printed circuit boards (PCB), presents an interesting problem. In this case, the electromagnetic fields are generated by the distribution of charges and currents on its paths and elements. In the near field region, the non-radiating reactive fields bring detailed information about the signature of the PCB under investigation. Accurate measurements of these fields can provide information about quality of the printed board and offer a number of advantages over methods utilizing electrical contact.

In the past, most theoretical and experimental work with respect to loop and spiral antennas focussed on the radiating near field region because of interest in electromagnetic interference effects [2].

The theoretical background for the solution of loop antennas and their derivatives results from the Maxwell-Faraday equation. Integrating this equation over the loop area S_a and applying Stokes Theorem [1], one obtains

$$\oint \vec{E} \cdot d\vec{l} = -j\omega\mu_0 \iint_{S_a} \vec{H} \cdot d\vec{S} \quad (1)$$

Depending on the geometry of the problem, this equation must be specialized to its particularities. Mathematical theory of a loaded circular loop in free space has been published [3], [4]. However, the classical loop antenna theory does not describe the printed loop or the printed spiral to the full extent.

2. Formulation

Consider a printed rectangular spiral antenna-sensor whose top view is depicted in Fig. 1. The cross-section of the device is shown in Fig. 2. The sensor is exposed to the reactive (or equivalently fringing) electric and magnetic near-fields surrounding the PCB. The portion of the near-field coinciding with the fringing field is commonly assumed to exist in the space confined at the distance

$1/k = \lambda/2\pi$ from an equivalent radiator. The sensor is placed inside a structure which consists of two parallel metallic planes of large dimensions. The distance between

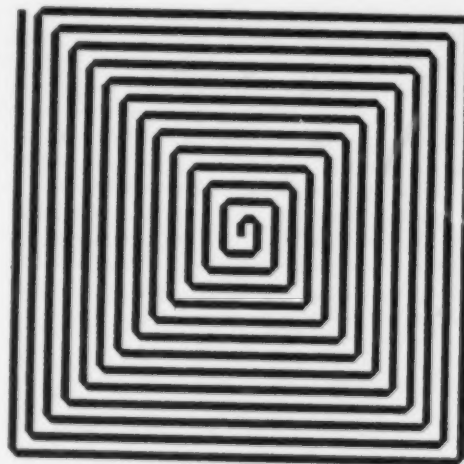


Fig. 1
Topology of the printed spiral antenna

the planes is small as compared with the wavelength of the upper frequency limit of the applicator. The purpose of this structure is to provide the eddy current shielding of the interior of the applicator. The Green's function of structure is completely determined by for this geometry of the contactless tester.

The fringing near-field performance of a single antenna sensor is of interest in response to the standard fields produced by the chosen radiators.

3. Method of Solution

A printed loop or spiral antenna may be modeled as an uniformly impedance-loaded loop. The boundary conditions satisfied on the metallic surface S of the sensor are [5]:

$$\vec{n} \times (\vec{E}^+ - \vec{E}^-) = 0 \quad (2)$$

and

$$\vec{n} \times (\vec{H}^+ - \vec{H}^-) = \mathcal{Y}_s \vec{E}_t \quad (3)$$

on the dielectric surface S_d . $\mathcal{Y}_s$ is the surface admittance of the dielectric layer given by

$$\mathcal{Y}_s = j\omega(\epsilon - \epsilon_0)d \quad (4)$$

where ϵ_0 is the substrate permittivity while ϵ_0 Equation (1) allows to find a general integral equation for the zero phase-sequence current by the use of the generalized Ohm's law

$$j\omega\mu_0 \int_{S_a} \nabla \times \int_{S'} G(\vec{x}, \vec{x}') \vec{K}(\vec{x}') \cdot d\vec{S}' d\vec{S}_a = \oint_L Z_L(\vec{l}) \vec{I}_0(\vec{l}) \cdot d\vec{l} - j\omega\mu_0 \oint_L \oint_L G(\vec{x}, \vec{x}') \vec{I}_0(\vec{l}') \cdot d\vec{l}' d\vec{l} \quad (5)$$

The second term on the right hand side represents interaction due to reradiated field. The contour integral in (5) is defined for a single loop. Therefore in order to apply (5) to the printed spiral antenna, segmentation of a spiral into elementary loops is imposed. Then for each loop-segment equation (5) is valid. The continuity of current $\vec{I}_0$ is applied at the separation of each constitutive loop.

Equation (5) is not tractable yet for numerical solution due to the fact both $\vec{I}_0(\vec{l})$ and $\vec{K}(\vec{x})$ are unknown. The Lorentz reciprocity theorem applied to the fields inside the antenna gives second condition which leads to the integral equation:

$$\begin{aligned} \int_{S_a} \vec{E} \cdot \mathcal{Y}_s \vec{E}_t dS + \int_{S_m} \vec{E} \cdot \vec{I}_0(\vec{l}) dS = \\ = \int_{V_a} \vec{I}_0(\vec{l}) \cdot \vec{E} dv \end{aligned} \quad (6)$$

where $\vec{E}$ is a function of current $\vec{K}(\vec{x})$ on the printed board. Equations (5) and (6) are coupled integral equations for the unknown currents. In order to solve the problem, it is natural to use the rectangular coordinate system because of the geometry in Fig. 1. The development of a method of solution of these integral equations and numerical results will be presented during presentation.

4. Experimental Results

Experimental arrangement is depicted in Fig. 2. Performance of printed spiral antennas was tested. The near field of a short dipole antenna was investigated by a sensor shown in Fig. 3 with separation distance 30 mm. The full 2-port scattering parameters were measured using HP8510B Automated Network Analyzer. Frequency band between 45 MHz - 1.0 GHz was of interest. It is evident from Fig. 4 and Fig. 5 that this type of sensors provide measurable signal (about 40 dB above noise level).

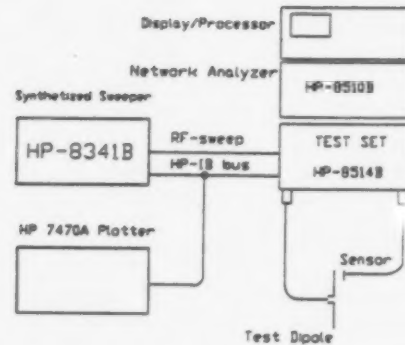


Fig. 2

Experimental arrangement.



Fig. 3

Photograph of the sensor.

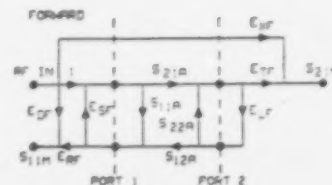
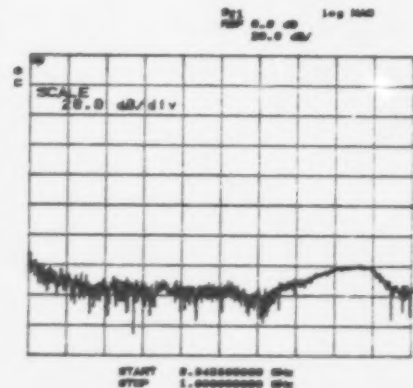


Fig. 4

Frequency Response of the S_{21A} coefficient
Forward Transmission Coefficient

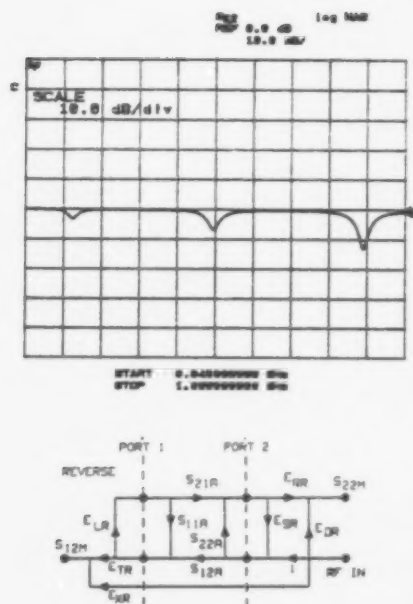


Fig. 5

Frequency response of the S_{22A} coefficient
Reverse Reflection Coefficient

5. Discussion and Conclusions

The candidate probe for the near-field imaging should satisfy several conflicting requirements:

- Small size;
- Flat frequency response over a wide range;
- High signal-to-noise ratio;
- Sensitivity both to electric and magnetic fields.

In order to reduce signal-to-noise ratio, the protective eddy current shielding is required. The broadband frequency performance of the sensor ensures that it is a good candidate probe for the reactive near-field imaging.

The theoretical analysis is essential to optimize the sensor response. Optimization of the sensor is the main purpose of the theoretical research to be completed.

6. References

- [1] EMI Noise Sensor, Noise Laboratory Co., Ltd., 2662-1, Noborito, Tama-Ku, Kawasaki City, Kanagawa Pref, 214 Japan.
- [2] United States Patent no. 4319248, "Integrated Spiral Antenna-Detector Device".
- [3] H. Whiteside and R.W.P. King, "The Loop Antenna as a Probe", IEEE Trans. on Ant. and Prop., AP-12, 1964, pp. 291-297.
- [4] M. Kanda, "Near Field Sensor for Simultaneous Measurement of Magnetic and Electric Fields", IEEE Trans. on Electr. Compat., EMC-26, No. 3, August 1984.
- [5] Y.N. Fel'd, "Diffraction of Electromagnetic Waves at Surfaces with Inhomogeneous Admittance", Radioelectronics and Eng. Physics, V.79, No.1, pp.1-8.

**ELECTRONIC SYSTEMS IN SEARCH AND RESCUE/
L'ÉLECTRONIQUE DANS LES OPÉRATIONS DE
RECHERCHE ET DE SAUVETAGE**

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

A COMPUTER-INTEGRATED RADAR SYSTEM FOR RESEARCH USE

D. Hamburger, C. Krasnor, S. Haykin, B. Currie, T. Nohara

Communications Research Laboratory, McMaster University,
1280 Main Street West,
Hamilton, Ontario, Canada, L8S 4K1

ABSTRACT

The Communications Research Laboratory (CRL) at McMaster University, has developed a coherent, dual-polarized radar system designed specifically as an instrumentation quality radar for research use. Significant computing capability is integrated into the radar system, with all radar functions, antenna movement and data acquisition under software control. Radar return data are stored digitally into computer memory in real time. This paper focuses on the computer aspects of this highly sophisticated system.

INTRODUCTION

Experimental programs involving the CRL in 1983 [1] and 1984 [2] using a land-based radar site in the Canadian Arctic clearly showed that the use of dual linear polarization significantly improves the ability to detect and classify multiyear and glacial ice. Those experiments also indicated the need for a highly versatile radar system that could be used to thoroughly study the effects of individual radar parameters and features.

To accurately control the many desired radar parameters without limiting the flexibility or expandability of the new system, a computer-based control system was required. It was also essential to have the radar return data available digitally, for statistical analysis [3] and signal processing [4].

The design and construction of this new radar system for research use began in 1985. The facility, called IPIX radar, has been successfully tested at a Canadian Coast Guard site at Cape Bonavista, Newfoundland. IPIX radar is at the heart of a major research programme aimed at improved detection of small targets in the presence of clutter.

GOALS OF IPIX RADAR

The main objective of the IPIX radar system is to collect data for research on various radar topics. In order to do this effectively, the IPIX radar must be flexible enough to study the effect of any single radar feature, or combination of different radar features. It must also be sufficiently modular to accommodate new features as they become available. The state of the radar system must be documentable and repeatable. The radar return data must be collected in a form that allows in-field processing. Moreover, data must be easily stored and recovered in its original form. The integration of computer-control into the design of the radar system is therefore necessary to provide for these demanding requirements.

DESIGN RATIONALE

The radar control system is set up with data registers external to the control computer, which can be read or written to through an interface bus. The radar's operation is based on the values stored in these registers. This method has several advantages. The full system status can be determined by reading all the registers. Any register (hence any radar parameter) can be changed independently of any other. The control computer only changes the state of the radar system, so that the computer itself is not required once the desired state has been established. This reduces the overhead of the control computer, freeing it for other tasks such as fault monitoring.

The computer aspects of the system are machine-independent. Programs are portable, and modular, so that the device-dependant sections can be replaced. This allows IPIX radar to accommodate advancements in computer technology (especially processing speed) with minimal

modifications.

A digital data acquisition system saves the radar return data only within a specified window of range and azimuth. This allows data memory to be economically used by storing only data from a selected area of interest.

DESCRIPTION OF THE RADAR CONTROL SYSTEM

A detailed description of the IPIX radar features can be found in earlier publications [5]. Here, only the computer aspects of the system are described.

The control interface for most operations is a simple parallel bus using TTL signals that can be implemented with four 8-bit parallel ports. The radar system status can be read, displayed, or saved by reading the values of the external registers on the parallel bus. All control software is written in the C programming language. Currently the control system is run on an IBM AT compatible computer. The radar features under computer control include transmit polarization, antenna position, sensitivity time control, pulse width and pulse repetition frequency. The computer also controls RF switches for selecting pulse compression and enabling the built in calibration equipment (BICE).

Transmit Polarization

The IPIX radar has two identical receivers, one connected to the horizontal channel of the dual-polarized antenna, and the other to the vertical channel. Both polarizations are received simultaneously. The transmitted signal is routed to either the horizontal or vertical antenna channel using a high speed ferrite waveguide switch. This switch also allows the transmit signal to alternate between the two channels on a pulse-to-pulse basis.

Antenna Positioning

The antenna is mounted on a two-axis pedestal capable of repeating absolute fixed positions to better than 0.1 degree accuracy. Rotation rate is also highly repeatable with spin rates up to 30 RPM. The antenna positioning unit uses a velocity feedback control loop to achieve a constant rotation rate with low drift (long-term variations) and low jitter (short-term variations). The antenna positioner maintains the pedestal position (azimuth and elevation) as binary encoded parallel TTL data with 16 bits of precision. The digital azimuth data is also available to the sampling system.

Sensitivity Time Control (STC)

The STC hardware varies the gain of the receiver as a function of time to remove range dependence of the radar return amplitude. Gain versus time functions are achieved according to STC curves. The STC curves are stored as digital data points in RAM (external to the control computer). After each radar trigger, STC curve datapoints are stepped through and D/A converted to provide control voltages which regulate receiver gain over time. The control computer has read/write access to the STC RAM. The STC curves can be loaded, edited, or copied to disk. This capability makes it possible to consider an adaptive STC system where the computer could modify the STC curves in response to different types of radar clutter (for example, due to varying sea states).

**Pulse Width and
Pulse Repetition Frequency (PRF)**

All timing pulses and periodic signals in the radar system are derived from a central highly-stable crystal oscillator. Digital frequency division makes other timing rates

available. Pulse width and PRF are selected from these values.

Switches

Remote-controlled coaxial switches are used extensively to allow the computer to control the RF signal path through the radar system. This permits, for example, the pulse compression system to be inserted or removed from use under automated control. These switches can configure the system for normal operation or for self-calibration (in which low-level transmit signals are routed directly into the receivers). Control of this sort can be used to automatically perform a calibration operation just prior to a data collection session. There are also four manually operated waveguide switches in the system. They are equipped with electrical position indicators with outputs available through the parallel bus. All switch positions are monitored by a hardware circuit that disables the transmitter if the switches are in an unsafe configuration.

DESCRIPTION OF THE DATA ACQUISITION SYSTEM

The data acquisition system is made up of two hardware components: the sampling system (which includes the A/D converters) and the data acquisition computer.

Sampling System

The radar analog output is in the form of four signals: I (in-phase) and Q (quadrature) signals for the horizontal channel, and I and Q for the vertical channel. After each radar pulse, return data are digitized to 8 bits per channel which are stored at the digitization rate (currently 30 MHz x 32-bits with a maximum of 50 MHz being possible with the existing hardware) into a sweep buffer. Data outside the window of interest (defined by range and azimuth as set by the control system) is ignored. Each sweep is tagged with a header that contains system information such as azimuth, polarization and the number of elements in the sweep. The data are then transferred at a rate of 2.5 MHz x 32-bits via a direct memory access (DMA) channel into a VMEbus memory card (currently 16 Megabytes).

Data Acquisition Computer

The memory card is part of the 68020 based data acquisition computer. Data for the selected number of sweeps are accumulated sequentially into the VMEbus memory. There the data are accessible for validation, processing or storage. The data acquisition computer operates under the UNIX operating system with software written in the C programming language and the Bourne shell.

Mass Data Storage

The 16 Megabytes of data on the memory card can be stored on digital tape in approximately three minutes. A high-capacity (2.2 Gigabytes per cassette) digital tape drive allows large numbers of data sets to be stored inexpensively.

TESTING AND USE

The IPIX radar has software diagnostic tools for validating the operation of its components and the quality of collected data. The data acquisition system can be tested extensively using external (hardware) and built-in data simulation sources. This process verifies that both hardware and software are functioning properly before actual data collection begins.

Once a data set is collected, diagnostic software may be used to check its format and general content. Statistical routines give immediate feedback on the specific content of the data. Because the data are held in RAM, these routines run quickly and can be applied to every data set collected. Graphic display routines are used to examine real data sets when operating the radar.

The radar equipment is tested and verified frequently during use. The data acquisition system is an excellent tool not only for collecting data, but also for monitoring the

performance of the radar itself. Conventional laboratory tools are also used to confirm the proper operation of the radar equipment.

The most recent field trip to the test site at Cape Bonavista, Newfoundland was in June 1989. Four hundred data sets, 16 Mbytes each, were collected and stored on tape during that trip. A data set could be collected, validated, and stored on tape in less than five minutes. Any data set can be restored from tape directly to RAM in its original form for further study and analysis.

BENEFITS OF COMPUTER-INTEGRATED RADAR

The immediate benefit of the IPIX computer-integrated radar system is that data sets can now be collected quickly and efficiently in a form suitable for processing. The system also provides flexible and repeatable control of radar parameters.

Ongoing research at the CRL has confirmed that signal processing [4], and statistical analysis [3] of radar return data improves ability to detect small targets in sea clutter. Many signal processing techniques require long time series data which is available through IPIX because of its unique windowing capability. Techniques of identification and classification of targets can now be investigated and implemented in a real time system.

Radar return data collected by the system is immediately available for signal processing and analysis. This makes adaptive radar feasible. Adaptivity can be achieved by allowing the radar system to alter its parameters automatically in response to real time analysis of radar return data. Specific areas of interest (large amplitudes of return, or recurring signals) found during wide area surveillance could be more thoroughly investigated by adjustment of radar parameters. Results of signal processing could be used to identify targets, to invoke other identification techniques or to modify the radar parameters further. For example, the system could point the antenna at a suspected target and collect long time series data. Once a target has been investigated and reported on, the radar could continue wide area surveillance. With an increasing number of software tools, analysis techniques, and radar hardware features, it becomes evident that an intelligent radar feedback/control software system is required. The hardware base and lower level software blocks already exist, making adaptive radar an achievable goal.

FUTURE PLANS

The IPIX radar has grown significantly beyond the objectives of the early designs of the system. The hardware modularity, computer integration, and digital processing capabilities of the system now provide the platform for innovative work in areas such as radar vision [6] and adaptive radar. Enhancements to the system are underway to further support these new research efforts.

Currently, the radar control functions are handled by the control computer, and data acquisition is handled by a separate computer. These two systems will be merged onto a single computer. This task is achievable because of the machine-independent design concepts of the system, and the portability of the C programming language. The system will have a single computer with access to the collected data, and the capability to control the radar based on evaluation of that data.

Graphics enhancements are planned to allow real time digital A-scope and plan-position indicator (PPI) displays. In keeping with the modularity and portability goals of the system, these graphics systems will utilize MIT's X-Windows device-independent graphics and windowing system.

Practical experience gained in the field has indicated areas where data collection can be streamlined. Record-keeping of sea, wind and weather conditions in conjunction with

gathering large volumes of radar data is currently being automated.

The computer control system provides excellent manipulation of radar parameters, but "experiments" generally involve tedious small changes to these parameters between consecutive data set collections. A higher level of radar control is now required, which specifies a group of data sets to be collected, rather than a single set. With the merge of the control and data acquisition computers (discussed above) this task becomes feasible, with only high-level programming required to complete the system.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support given to the IPIX project by Litton Systems Canada Limited, the Natural Sciences and Engineering Council, the Ontario Ministry of Colleges and Universities, and the Department of Fisheries and Oceans.

REFERENCES

- [1] Haykin, S., Currie, B.W., Lewis, E.O., and Nickerson, K.A., 1985, "Surface-Based Radar Imaging of Sea Ice", *Proc. IEEE*, Vol. 73, No. 2, 233-251.
- [2] Lewis, E.O., Currie, B.W., and Haykin, S., 1985, "Effect of Polarization on the Marine Radar Detection of Icebergs", *IEEE International Radar Conference (Radar 85)*, Washington, D.C.
- [3] Nohara, T., Haykin, S., Currie, B.W., and Krasnor, C., 1989, "Towards the Improved Detection of Small Ice Targets in K-Distributed Sea Clutter", 1989 International Symposium on Noise and Clutter Rejection in Radars and Imaging Sensors, Nov. 1989, Kyoto, Japan.
- [4] Krasnor, C., Haykin, S., Nohara, T., and Currie, B.W., "A New Coherent Marine Radar for Ocean-Related Studies", *Canadian Conference on Electrical and Computer Engineering*, 1988, Vancouver, Canada.
- [5] Krasnor, C., Haykin, S., Currie, B.W., and Nohara, T., "A Coherent Dual-Polarized Radar for Ice Surveillance Studies", *International Conference on Radar*, 1989, Paris, France.
- [6] Haykin, S., "Radar Vision", *Canadian Conference on Electrical and Computer Engineering*, Sept. 1989, Montreal, Canada.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

RADAR VISION

by

Simon Haykin
Communications Research Laboratory
McMaster University
Hamilton, Ontario

ABSTRACT

In this paper, we describe a novel idea for which we have coined the term "radar vision". The goal is to make radar into an intelligent remote-sensing device that is capable of developing an understanding or cognition of the surrounding environment. We also describe the facilities that are already in place at McMaster University for research into the development and perfection of a radar vision system.

1. INTRODUCTION

Radar has established itself as an indispensable tool for both civilian and military applications [1]. In this paper we discuss a novel idea, for which we have coined the term *radar vision*. The goal here is to make radar an *intelligent remote sensing device*, such that it is capable of developing an *understanding or cognition* of the surrounding environment. It is our strong belief that such a capability is indeed within our reach, given the recent advances in signal processing, symbolic processing, computer vision, neural networks, and parallel computing.

To illustrate the main points of this new idea, we use the case of a coherent marine radar for ocean surveillance. Nevertheless, the idea of radar vision as described herein applies to other radar environments (e.g., Arctic surveillance, air traffic control, weather monitoring, navigation along a confined waterway, etc.). To the best of our knowledge, radar vision as described in this paper has originated with this author.

The remainder of the paper is organized as follows. In Section 2, we briefly review the use of radar as an electromagnetic sensor. In Section 3, we view the environment as a *teacher* to radar. In Section 4, we discuss the similarities and differences between human vision and radar. In Section 5, we outline the notion of *radar vision* that may well represent the way of the future for radar design. In Section 6, we briefly describe the facilities for radar vision research at the Communications Research Laboratory, McMaster University; the ultimate aim of the research is to design a radar vision system for the ocean environment. We finish the discussion in Section 7 with some concluding remarks.

2. RADAR AS A SENSOR

Radar is an electromagnetic device in that it interacts with the surrounding environment by transmitting an electromagnetic signal. The received signal that it processes is made up of echoes produced as a result of reflections of the transmitted signal from various objects located in the environment. This interaction is achieved through the use of an antenna designed to operate over a frequency band of interest. Indeed, radar has evolved into a sensor that is essential to many facets of our highly technological society. Important applications of radar include its use as a sensor for weather monitoring and forecasting, air traffic control, navigation on the high seas, probing the subsurface, and many other applications.

Typically, radar constructs a *two-dimensional electromagnetic image*, with range and azimuth as the two polar coordinates of the image. Naturally, the details of the image depend on the application of interest. Moreover, the final image displayed for use by a human operator is a processed version of the raw image picked up by the radar antenna. In this final image, specific objects of interest (e.g., ships, aircraft, weather formations) are highlighted, whereas reflections from unwanted objects (e.g., ground, sea) are suppressed.

The number of independent images, and therefore the amount of information extracted by the radar, is determined by the dimensionality of the radar signal; this, in turn depends on the manner in which the available discriminants are employed. Basically, *electromagnetic (EM) fields*, and therefore radar, can be described by four discriminants:

1. **Time:** This discriminant arises due to the wave nature of the EM field, and the temporal properties of the environment in which the radar is embedded.
2. **Space:** This discriminant refers to the fact that EM fields exist in three-dimensional (3D) space.
3. **Frequency:** This discriminant is the dual of time; it re-emphasizes the fact that the EM field is a wave phenomenon.
4. **Polarization:** This final discriminant refers to the fact that the EM field is a vector field.

Even though these four discriminants may not be explicitly exploited in the design of a radar system, nevertheless, they are all involved in the operation of the system. For example, in a radar using a conventional (e.g., dish antenna), the physical aperture of the antenna represents the extent of the spatial discriminant and, hence, controls the beamwidth of the antenna. Indeed, in this paper we consider the case of a marine radar using a conventional antenna.

Moreover, these four discriminants may be viewed as the independent variables over which the dependent variable is measured. In radar, the dependent variable is represented by the transmitted signal or received signal, depending on the point of interest in the system. In the case of a *narrow-band radar* system (i.e., narrow band with respect to the carrier frequency), the radar measurements are fully represented by the 2-by-2 *polarization matrix*. If the radar is *coherent*, then the radar measurements at different points of time are related to each other. In such a radar, each element of the matrix has a *complex amplitude* (i.e., amplitude and phase). Moreover, the polarization phenomenon obeys reciprocity. Accordingly, the polarization matrix of a coherent radar is uniquely defined by 5 independent parameters, except for a phase factor.

Each radar parameter has a distinct physical significance of its own:

1. **Amplitude** varies with the visible size (i.e., radar cross-section) of a target.
2. **Coherence** makes it possible for the radar to respond to radial motion (i.e., Doppler) of a target.
3. **Frequency** enables the radar to tune to spectral characteristics (i.e., resonances) of a target. Moreover, through the use of *frequency diversity* (e.g., the simultaneous transmission of a fixed frequency and an agile frequency), it becomes possible to measure the size of a target [2].
4. **Polarization** enables the radar to respond to depolarizing characteristics of a target and, therefore, aid the recognition of the target.

By varying the latter two radar parameters, namely, frequency and polarization, the radar becomes *controllable* and therefore capable of *adapting to variations* in the surrounding environmental conditions.

In summary, then, radar is a *controllable, active sensor*.

3. THE ENVIRONMENT AS A TEACHER

Typically, the environment in which a radar system operates is continually changing. These changes are brought on

by two principal factors: (1) natural variations in weather, and (2) motion of targets relative to the radar. Accordingly, the received radar signal is in reality a sample function of a *nonstationary stochastic process*.

For example, in an ocean environment, that includes targets of interest, the composition of the received radar signal is influenced by the following factors:

1. Sea State, which is itself determined by the prevalent weather conditions.
2. Natural objects, such as ice targets, depending on geographic location and the season of the year. The ice targets may include icebergs, growlers (i.e., small fragments of ice broken off an iceberg), and pack ice.
3. Man-made objects, including ships (of varying sizes), fishing boats and trawlers.
4. Atmospheric propagation effects, such as ducting, multipath, etc.

The radar is typically slaved to monitor the surrounding environment on a continuous basis. Hence, the system possesses a *natural* capability of gathering large volumes of valuable information about short-term as well as long-term variations of the environment. Yet, because of storage considerations, radar processing is usually limited to one or two scans of data; that is, only short-term memory is used. Moreover, except for the radar targets (e.g., ships, icebergs) being tracked, radar echoes from other objects (e.g., sea surface) are usually viewed as a "nuisance". The term "clutter" is used to refer to such radar echoes. Despite the fact that radar clutter may actually contain valuable information (e.g., the orientation of the ocean waves, and their wavenumber spectrum), nevertheless, it is filtered out at early stages of processing. In so doing, valuable information about the surrounding environment that could be used to enhance the processing is thrown away.

The point we wish to stress is that the environment in which a radar system operates is a *natural teacher*. Over the course of time, the environment exposes the radar to a wide range of external conditions. The radar can become more experienced with time by learning from the environment. To learn from the environment, however, the radar would have to be equipped with not only *short-term memory* but *long-term memory* as well. The learning itself may be carried out either in a *supervised* or *self-organized* fashion. In the context of radar, self-organized learning has a great deal of appeal, as it would enable the radar to learn how to characterize the received signal in *real-time*. Whatever form of learning is used, clearly, the radar must go through a *training process* where it has direct knowledge of the input signals and feedback about the correctness of interpretations made.

4. SIMILARITIES AND DIFFERENCES BETWEEN HUMAN VISION AND RADAR

There are several important similarities between human vision and radar to warrant research into a new area for which we have coined the term "Radar Vision". Naturally, there are basic differences between them that set them apart.

The similarities between them are:

1. Human vision operates by light waves, whereas radar operates by electromagnetic waves.
2. Receptors in human vision are represented by eyes, and by an antenna in radar.
3. The processing of signals is performed by the brain in human vision, and by a combination of analogue and digital circuits in radar.
4. Human vision and radar are both controllable.
5. In human vision the output is the perception of a three-dimensional (3D) image of a scene, whereas in radar the output is typically a two-dimensional (2D) display of the surrounding environment. It is important to note, however, that radar is not necessarily a 2D system. Furthermore, in radar the *depth information* (represented by the range of a target) is measured directly, whereas in human vision the system must rely on *depth cues*. That is to say, radar images depth information, whereas the human perceives it.

As for differences between human vision and radar (in its present form), we may identify the following:

1. Human vision possesses a highly sophisticated learning and adaptive capability. By contrast, a radar system employs a highly primitive form of adaptation, if at all.
2. In the context of "focus of attention", human vision exhibits an extraordinary ability to provide global and local focus. A typical radar, on the other hand, is limited to a global or a local focus of attention, but not both.
3. Human vision can handle the processing of a massive volume of data, which may only be equalled by a combination of as many as 1000 supercomputers. On the other hand, the volume of data generated and processed by a fairly sophisticated radar system (e.g., weather radar or air traffic control radar) is large by ordinary standards, but minute by comparison with human vision.
4. Human vision relies on external illumination for its operation, whereas radar supplies its own illumination. Moreover, unlike human vision, radar sees through fog and cloud.
5. Spatial resolution in radar is much poorer than in human vision. For a specified wavelength, spatial resolution in radar is limited by the physical size of the antenna aperture.
6. Unlike radar, human vision is both phase and polarization blind.

This brief discussion suggests that much can be learned from human vision and transplanted into design of a radar system. Indeed, the use of radar vision, described next, is intended to rectify the deficiencies of present radar designs. A radar facility is often quite expensive as, for example, in weather monitoring or ocean surveillance for ice management. Through the use of radar vision, it should be possible to make an expensive facility that much more valuable!

5. RADAR VISION

From analogy with computer vision [3], we may identify a hierarchy for radar vision that consists of three stages, as indicated in the block diagram of Fig. 1. The first stage, *low-level radar vision*, is designed to perform preprocessing on the incoming radar data. The second stage, *intermediate-level radar vision*, performs target tracking and feature extraction. The final stage, *high-level radar vision*, performs knowledge processing for the purpose of target interpretation. Typically, the raw radar data at the input of the system in Fig. 1 contain 10 million pixels per scan. The image computed at the output of the system, on the other hand, may be described by about 0.3 bits per pixel.

The *feedback* link shown in Fig. 1, from the high-level output back to the radar, plays a key role in radar vision. Specifically, it enables the system to develop an *adaptive control* of the environmental illumination [2]. If the high-level radar vision is confused by what it "sees" through the radar, the control mechanism reconfigures (adapts) the radar to provide a different "look" at the target. This *interactive probing* of the target and its environment continues (on the order of seconds) until recognition takes place.

5.1 Low-level Radar Vision. The function of this first stage is three-fold

- (a) To perform noise removal, and contrast enhancement.
- (b) To fuse data collected by the like- and cross-polarized channels of the radar (if applicable) into a *composite complex-valued image*. This image is typically in polar form.
- (c) To segment the composite radar image into distinct regions and assign identifiable characteristics to each region. A region may consist of one or more radar resolution cells; by a "resolution cell" we mean a cell that is uniquely defined in range by the transmitted pulse duration and in azimuth by the antenna beamwidth.

Low-level radar vision is an attractive application of neural networks [4,5], because it involves the use of image processing algorithms that are essentially *local* in the 2D topology of the image. The processing of one pixel of the image depends only on neighbouring pixels.

5.2 Intermediate-level radar vision. In ocean surveillance as encountered on the East Coast of Canada, for example, the radar objects of interest include icebergs, small fragments of ice (i.e. growlers), and fishing boats. The function of this second level of radar vision is to track the movement of icebergs, growlers and fishing boats. The tracking of icebergs is relatively straight forward because of their large radar cross-section. The tracking of growlers, however, is more difficult because of their small radar cross-section and their occasional obscuration by ocean waves. A strategy well-suited to the situation described herein is to *track and then detect*; this is in the opposite order to traditional radar practice. Such a strategy may give rise to *false tracks* due to sea spikes, and *interruptions* in target tracks due to target obscuration by ocean waves. Through a learning process of radar tracks from one scan to the next, it should be possible to eliminate false tracks and to bridge gaps in target tracks. A real challenge for the learning process is to discriminate between two target tracks that merge with each other tangentially for a good part of the way or cross each other in highly interrupted parts of the target tracks.

5.3 High-level radar vision. This level of radar vision involves *knowledge processing*. It has two inputs: (a) information extracted from the intermediate-level radar vision, and (b) knowledge of the environment gained by a human observer or other means. The function of this level is to combine these sources of knowledge so as to compute a *dynamic three-dimensional (3D) image* that exhibits *cognition* or *understanding* of the ocean scene, with objects of interest in the environment identified individually. Specifically, the high-level radar vision has a two-fold function:

- (a) To provide a 3D computer visualization of the ocean surface surrounding the radar, and to identify those parts of the surface covered by pack ice (if present).
- (b) To *discriminate* between icebergs, growlers and fishing boats (unequipped with transponders for active communication with home base).

Successful discrimination between growlers and fishing boats would have a significant impact on *search-and-rescue* operations in an ocean environment.

5.4 Knowledge of the Environment. High-level radar vision has the function of not only organizing and controlling the flow of information from the intermediate-level radar vision but also combining it with knowledge of the environment. The overall motivation here is to compute a 3D radar image of the surrounding environment. In ocean surveillance, knowledge of the environment used as an external input to the high-level radar vision is as follows:

1. The (nearly) parallel streaks seen on a radar image represent the peaks of ocean waves for the case of well-structured sea swells; the *Hough transform* may be used to enhance the visibility of these lines. On the other hand, when the sea is *choppy*, the corresponding radar image is random in appearance. Spectrum analysis of radar data along range rings may provide additional information on the wavenumber spectrum of the ocean surface. Putting all of this information together, and combining it with information on wind direction and sea state, we may compute a 3D image of the ocean surface.
2. *Radar shadows* may be used to identify targets with significant vertical relief.
3. The initiation of *false tracks* is most probably due to sea spikes that occur randomly.
4. Interruptions in target tracks are most probably due to *target obscuration* by ocean waves. Fishing boats usually have radar reflectors mounted on a mast as high as possible; hence, unlike growlers, they are rarely hidden by ocean waves in low to medium sea states. In such sea conditions, radar targets whose tracks exhibit frequent interruptions may therefore be inferred to be growlers rather than fishing boats.
5. Passive floating objects move under the control of wind conditions and ocean currents while motorized boats tend to move along pre-destined paths. Moored buoys and

grounded objects will exhibit still other distinct movement patterns. Thus, tracing the movement history of a target and comparing it with that of ocean currents and wind flow may lead to its identification.

6. The surface of a pack ice exhibits a wave-like structure that is influenced by, and yet distinctively different from, ocean waves in open water.
7. Knowledge about the depolarizing characteristics of ice targets may aid their recognition.
8. Other high-level knowledge of the environment, obtained by human observations.

5.5 Adaptive Control of Radar Observations. This capability makes it possible to readily adjust the radar illumination of the surrounding ocean environment; see the feedback link of Fig. 1. For example, the transmit radar frequencies and polarizations may be varied so as to respond to statistical variations of the ocean environment and thereby minimize the effect of sea clutter. Another possibility is to switch from *scanning* to *staring mode* of operation and focus processing power on a window centred on a target of small radar cross section (e.g., growler of fishing boats), and thereby enhance its detection. The form of adaptive control described herein lends itself to the use of the *backpropagation algorithm* and *neurocontrol* [6].

6. FACILITIES FOR RADAR VISION RESEARCH

At the Communications Research Laboratory (CRL), McMaster University, we have put together a research programme aimed at the development and perfection of a radar vision system, as described in the previous section. Central to this study is a highly sophisticated radar designed and built at the CRL with financial support provided by the Natural Sciences and Engineering Research Council (NSERC), Litton Systems Limited (Toronto), and the Provincial Government of Ontario. The radar is known as IPIX, which offers the following unique system features [7,9]:

1. Full *coherence*; hence, the capability to measure Doppler produced by target motion.
2. *Pulse-to-pulse polarization switching* (i.e., near simultaneous transmission of the two orthogonal polarizations, horizontal and vertical, and their simultaneous reception); hence, a capability to measure the *full polarization matrix* defined by five independent real-valued radar parameters.
3. Simultaneous transmission and reception of two frequencies, one fixed and the other *agile*; hence, the possible use of *frequency diversity*.
4. *Pulse compression*; hence, additional range coverage and improved resolution.
5. A unique method of *data collection* that encompasses the following attributes:
 - (a) Direct acquisition of radar data into computer memory in real time.
 - (b) Computer control of data selected for storage on the basis of azimuth and range; hence, the reduction of data volume to be stored to manageable amounts.
 - (c) Ability of the computer to "see" the data; hence, such matters as data validation, processing, reduction, etc., can be performed before the data are stored on tape.
6. *Adaptive operation* of the system by virtue of the "acquisition-control" loop built into the design of the system. This will allow for modification of the radar parameters to suit the target being examined at every instant of time.
7. Ability to operate the radar in a *scanning* or *staring* mode. In a scanning mode, the dual-polarized radar antenna is rotated uniformly at an adjustable speed. In a staring mode, on the other hand, the radar antenna is permitted to dwell along a "look direction" of interest for a duration of time controlled by the computer.

IPIX has been *successfully tested* at a Canadian Coast Guard site at Cape Bonaville (Newfoundland). The results obtained thus far have been most encouraging. In particular, we have

discovered a Doppler signature, namely *winking* that is associated with small objects (e.g., small pieces of ice and waverider buoys) floating in open water, a feature not known hitherto. It is intended to exploit this signature for the enhanced detection of weak targets in clutter.

Given the full capability of IPIX, we are able to form 10 independent images of the environment. Clearly, it is impossible to display all of these images to an operator. Provision will therefore be made for their optimum *fusion* into one colour-encoded image in real time.

The software for much of the computer processing and data fusion is being implemented on a *Warp systolic machine*. This highly parallel processor was conceived and developed at Carnegie-Mellon University for vision research (applied to robot control). The machine was purchased under funding from the Telecommunications Research Institute of Ontario (TRIO); it is integrated into a sophisticated signal processing computer that incorporates Sun workstations, a colour Data Cube display, an APTEC I/O computer, a VAX 8300 computer, and an interface into analogue and digital radar recording facilities.

Thus, with the combination of IPIX and Warp, we, at the CRL (McMaster), are well-poised for undertaking the research needed to perfect radar vision. The results of this research will be presented in future publications.

7. CONCLUDING REMARKS

The extraction of lines, edges and shapes, and the recognition of objects (targets) of interest are all natural operations for human vision. These operations, on the other hand, are rather difficult for present-day computers to perform. However, through advances being made in computer vision and related areas, we are beginning to develop answers to these problems.

As for the capability of a present-day radar for vision, even though it may be highly sophisticated, at best we may compare it to that of a one-year old baby. Nevertheless, in view of the continuing advances being made in signal processing, symbolic processing, neural networks, computer vision, and parallel computers, the idea of radar vision is indeed a practical feasibility.

It should also be emphasized that although the main points of radar vision, as described herein, have been made in

the context of a marine radar for ocean surveillance, nevertheless, the idea of radar vision is basic to all remote sensing applications of radar. Moreover, in a phased-array radar system that uses an array of radiating elements as the antenna, the system is able to perform *spatial processing* of the received signal in addition to temporal processing. Accordingly, a phased-array radar system has the potential for radar vision at a more sophisticated level than that described in this paper.

ACKNOWLEDGEMENTS

The author is indebted to the Natural Sciences and Engineering Research Council (NSERC) for supporting the research described herein. He is also grateful to many of his colleagues at the Communications Research Laboratory, McMaster University, for their useful inputs that have helped shape the paper into its present form.

REFERENCES

- [1] Skolnik, M.I., "Introduction to Radar Systems", Second Edition, (McGraw-Hill, 1980).
- [2] Gjessing, D.T., "Adaptive Matched Illumination Radar", (IEE London), 1986.
- [3] Fischer, M., and O. Firschein, "Intelligence: The Eye, The Brain, and The Computer", (Addison-Wesley, 1987).
- [4] Bienenstock, E., in "Neural Computers", pp. 61-70 (Springer-Verlag, 1988).
- [5] Anderson, J.A., "Mathematical Biosciences", vol. 14, 1972.
- [6] Werbos, P.J., "Backpropagation and Neurocontrol: A Review and Prospectus", International Joint Conference on Neural Networks, Washington, D.C., June 1989.
- [7] Krasnor, C., S. Haykin, T. Nohara and B. Currie, Canadian Conference on Elect. & Comp. Eng., Vancouver, November 1988 (invited).
- [8] Krasnor, C., S. Haykin, B. Currie and T. Nohara, Int. Conf. on Radar, Paris, April 1989.
- [9] Hamburger, D., C. Krasnor, S. Haykin, B. Currie, and T. Nohara, "A Computer-integrated Radar System for Research Use", Canadian Electrical and Computer Engineering Conference, Montreal, September 1989.

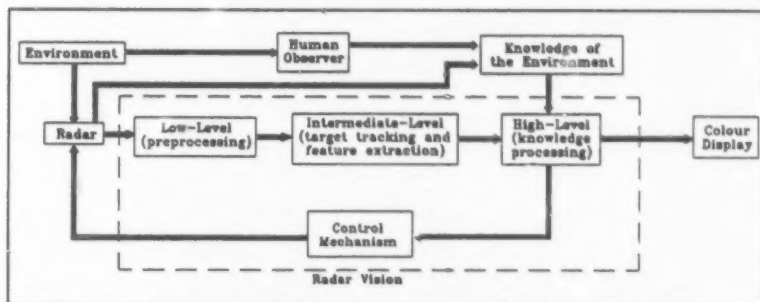


Figure 1. Block Diagram of Radar Vision System

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

CANSARP: A COMPUTERIZED SEARCH AND RESCUE PLANNING
PROGRAM FOR THE CANADIAN COAST GUARD

Byron C. Berglund and Donald O. Hodgins

Seaconsult Marine Research Ltd.
820-1200 West 73rd Avenue
Vancouver, B.C., Canada V6P 6G5

Abstract

The planning and coordination of an extensive ocean search involves the management and manipulation of tremendous amounts of information in realtime. Although some information used in search planning is relatively static, for example current and tidal atlases, most critical information is dynamic and unpredictable. The search planner must coordinate information from many different sources, first to define the incident characteristics and plot a plausible drift track for the search target, then to assign areas to the search units at his disposal. A new version of the CANSARP program is described, specifically dealing with the information load on search planners, including the data structures necessary to model the information typical in large scale ocean search missions.

Coastal Search Planning in Canadian Waters

Over the past 10 years the advantages of using PC computers to aid marine search and rescue planning have been recognized in Canada and many other nations. In 1985 the Canadian Search and Rescue Planning program (CANSARP) was developed, and became operational in most rescue centres within two years.

Our goal when we started CANSARP V1 in 1985 was to automate the search procedures as they are defined in the oceanic search planning manual. The accepted procedures are chosen mainly for their ease of application by hand calculation and for maximizing the probability of locating the search target. Seaconsult [1] focussed attention on using oceanographic expertise to provide better estimates for the currents; the procedures used to sum the components of the drift and arrive at a search area were still exactly as laid out for hand computations.

In developing CANSARP V2, it was recognized that calculating the drift track and the search area is only the beginning of the search procedure. Demands were made for the computer to help with the management of the sometimes erroneous and always changable information about the weather conditions and even the search target. An attempt was made to automate the allocation of search units to the computed search area, again following the accepted manual procedures. However, the search manual concentrated on procedure and the need for information management was not given as much emphasis. This version identified the shortcoming of CANSARP in the user interface and management of information; the input to the program was through a linear sequence of menus driven from the function keypad.

In 1988, another program, DRIFTALC [2], was written specifically to handle the tidal flows dominant in the Straits of Georgia and Juan de Fuca and the unique features of search planning in a region of islands and channels. This program had a much better user interface than the previous CANSARP programs; it featured a mouse-driven input system running on an AT class computer.

CANSARP Version 3

CANSARP V3 will be focussed initially on the user interface in an attempt to match the power of the computer with the needs of the search planner. When an incident is first reported, the information is often sketchy and in a lot of cases assumed. During the course of the search, new information becomes available in a unpredictable manner, and it is very difficult for search planners to deal with changing information while directing the search. The design of CANSARP will correspondingly change in direction to focus on the management of information, with calculations based on the available information taking a less important, although still significant, role.

Search Mission Information

The information which the search planner must deal with includes several independent incidents, each of which might have dozens of scenarios, some active and some which might have been rejected. Every stage of the search plan must be documented, including the reasoning for choosing and rejecting the different scenarios. Each scenario also includes the initial assumptions and calculations included in the drift, along with the weather conditions used to drive the calculation. The reasoning behind the assumptions should also be documented. Given a choice of active or likely scenarios, each will define an area with a computed radius, with a specific probability that the target is in fact there. These "areas of probability" must be searched. The search planner must divide the areas to be searched among the search units he has at his disposal, each of which has different characteristics and search efficiencies. In addition, areas which have been searched have their probabilities reduced by a factor which depends on both the weather and the search unit. As the weather conditions change, as new information becomes available, or as forecasts are replaced by observed conditions, the calculations which depend on this information must be redone and changes to the search datum noted. The search planner must be notified of serious changes to active searches. The information the search planner deals with in his day-to-day duties includes the relationships between facts as well as the facts themselves.

Search Planner Training

A well documented search plan serves both as legal justification and as a training tool for new search planners. At any stage of an active or completed search plan stored in the computer, a user may review the options available to him, and the choices made by the controlling search planner, documented with his reasons for making them. Simulated incidents can be derived from these stored searches and the information released in realtime fashion, so that trainees may gain experience and develop the skills required to direct an ocean search plan. This is one of the important benefits of automating the management of the information involved in a search mission.

Initial Incident Information

The initial information about an incident falls into four broad classes:

- (1) Last Known Position: based on a sighting or knowledge of where the vessel was meant to be.
- (2) Uncertain time: overdue vessels known to be in a particular area.
- (3) Uncertain position: overdue vessels taking a known or assumed route.
- (4) Debris: sighted or washed ashore, which must be reverse-drifted for an uncertain time.

Each of these four classes define an incident position when fixed at a particular time. This is the position from which the target must be drifted forward to its position at search time. In addition, there are uncertainties in the type of target, whether it is a person, a life raft, or a boat. Each piece of information has a probability attached which gives some measure of its worth. These initial probabilities are assigned by the search planner, based on his expertise, and are subject to change as the search progresses and new information becomes available.

The assumptions taken about this information form the basis for a scenario. A typical search involves several scenarios which define an area of probability; the greatest effort must be concentrated in the area with greatest likelihood. The search planner must be able to manipulate several scenarios simultaneously and weigh each against the others in probability.

The Scenario

A scenario evolves from the initial specifications of target type, leeway characteristics and location, through drifting to a target datum and associated area of probability, hence search area. During the drifting of an assumed target many more assumptions must be made concerning the weather and seastate conditions along the drift track. Again, these assumptions may be assigned probabilities, and the information often changes dramatically. For example, the assumed conditions may have wind at 30 knots from the north, which implies a southerly drift. Later, when search units enter the area, they report 50 knot winds from the southeast, implying a northwesterly drift. The search planner must be able to incorporate this new information immediately to see how it affects the search area in order to redirect the units on station. The flow of information must be automatically managed to reveal its effects in timely fashion.

With the computer handling the information and keeping track of the assumed probabilities, the search planner is freed to concentrate on defining the search areas and assigning search units. With a suitably powerful and intuitive interface between the user and the computer, search planners can then engage in "what-if" scenarios which may never in fact end up being used. This trial procedure allows the planner to hone his skills and improve his judgment about probabilities, and ultimately increase the chances of successful rescue.

The Scenario Data Structure

Each search mission has an associated set of scenarios. The scenarios are a sequence of positions at specified times, distinguished by the initial assumptions and calculations included in the drift, which in turn depend on the assumed weather conditions along the drift track. Several scenarios may share an initial sequence of positions, and they are then distinguished only from the point where they diverge.

Thus, each scenario is in fact a tree rooted at one of the four initial information classes, with internal nodes at points along the drift where conditions differ, and leaves at the datum locations (each with an associated area of probability). The edges connecting nodes represent the drift of the target from one position to a position later in time. A typical search may have only one incident report and initial information type, hence only one scenario tree. A number of rejected scenarios may also exist as entire trees are rejected because of faulty initial information, or as branches of active trees are pruned because of faulty information at the internal node where the tree branched.

For example, a search mission is started based on a report of an overdue fishing vessel known to be in a certain area (initial information type 2). Someone calls in and reports a sighting (type 1), which is later determined to be a case of mistaken identity. Later debris is picked up and identified as belonging to the missing vessel (type 4). Each of these scenario trees must be followed up, but the tree rooted at the mistaken sighting must be rejected, possibly after some effort has been spent searching. The tree of scenarios may be pruned during a search mission, but it can not be discarded entirely; the documentation of the case must be kept, and the branch may later be reactivated as new information comes in.

Nodes

The nodes in the scenario tree represent "fixed" target positions, in the sense that they are the basis for future drifts from this point. There are the initial position, points where assumptions about the weather conditions change, or the time and area where searches are conducted. Hence, the positions cannot be changed; new possible positions must be represented in a node further up the tree, i.e. earlier in time, and an entirely different subtree. Nodes represent points where multiple drift tracks diverge, each with a different set of assumptions. Each node in the tree will therefore have a document associated which explains the reasoning behind the divergence in assumptions. The different assumptions each mark the tail of a different edge.

Edges

Each edge in the scenario tree is directed, with its length being the time elapsed from the position at its tail to the position at its head. The edge contains a link to the associated weather data assumed to hold for the corresponding time and place, and associated calculations which, when applied for the length of time represented by the edge, imply a drift which takes the search target to the position at the edge's head. The target's position at any time represented within the edge's length is found by integrating the calculations over the time desired. Note that a simple scenario with unchanging conditions will consist of the root node (initial information) and a single edge. Earlier versions of CANSARP maintained just this simple case; branches in the scenario tree were represented in additional "search files," distinguished by slight changes in the search name, e.g. Molly 1, Molly 2, etc. Many of these files would duplicate initial information and drifts. The user was left to distinguish and maintain this multitude of files with no documentation aid from the computer. Every point along the drift specified by an edge is the position of a particular search target at an instant of time.

Using the System

The SeaWeather Data Editor

The basic information driving the CANSARP system are the assumed weather conditions, the derived surface current, and the tidal and geostrophic sea currents, where appropriate. A large part of the management task will therefore be maintaining the weather database. The Atmospheric Environment Service (AES) has recently made its observed and forecast condition reports available on a satellite network at low cost. However, this information are available only at selected locations along the coast, or from ocean-going vessels and weather buoys. They will therefore only form the basis of assumed conditions and often must be extrapolated to the scene of an incident. On-scene reports once search units are deployed also enter into this database and must be reconciled with the extrapolated results to obtain updated conditions. The system will therefore include a graphical weather data interface which will allow the user to include or specify conditions at a number of locations, extrapolate the conditions over an area, and view the result in usual weather chart form. As conditions are edited and the search progresses, calculations which depend on the data must be recomputed. This will create a new set of scenarios, branching from the old where data changes.

The PlaceNames Geographic Database

The search plan functions of CANSARP are performed on a geographical tableau representing the area of interest. The user can view this tableau on his screen and indicate points of interest with the mouse, either by simply pointing, or by watching the echoed longitude and latitude of the cursor position until the correct position is found. However, incident reports rarely are in terms of precise locations, typically a local name is mentioned instead. The search planner must be able to map these names to a physical location for specification to CANSARP, and for passing on to the search units in terms of longitude and latitude. Planners cannot be familiar with all names in their area of responsibility, thus the need for a PlaceNames utility.

Simply stated, the PlaceNames Database is a mapping of names to longitude-latitude positions. This mapping is many-to-one; a single position or area might have many names, each denoting a different scale. For example, the area around Brooks Peninsula could also be referred to on a larger scale as Vancouver Island. This mapping is also invertible, that is given a position, the corresponding names can be found. The inverse mapping is used by the CANSARP system to name areas displayed on the screen, with the choice of name dependent on the scale of the map being displayed.

Defining a Scenario

Beginning with an on-screen graphic of the coastline in his assigned area, for example the Pacific coast of Canada, the search planner must indicate that an incident report has been received. He may indicate the area by pointing with the mouse at a specific point along the coast or offshore. The position of the cursor is echoed in longitude and latitude so that the task of specifying an exact position is simplified. When a mouse button is pressed, a popup menu appears at the cursor and allows the user to indicate the report type, i.e. the type of the initial information, and to complete a form with the information.

Once the initial information has been defined, the search planner chooses an incident time and the position at this time is indicated on the screen. The user

has the option of using defaults throughout his interaction, so that reasonable responses from the computer are assured. For example, if the search planner initiates an incident and begins a drift to a position later in time without specifying anything but the initial information, the computer will make assumptions so that a reasonable position can be determined. The user is free of course to alter these assumptions at his convenience by being as specific as he believes is necessary. This automated default facility extends from extrapolating the AES weather reports extracted from the satellite transmission as a first estimate of the weather in the incident area, through to selecting calculations to be included in the drift based on the area. For example, the computer can determine the correct make-up of tidal, sea current, and wind-driven currents that should be included in the drift calculation.

Creating a Branch

Once the search planner has defined an incident and the initial scenario, whether by specific choices or by taking the defaults supplied by the computer, inevitably changes must be made. Changes to the initial information, for example the search target type, are made as for the first case, by indicating an area with the mouse and defining a new scenario. Changing the set of calculations used in the drift, for example using surface drifters to determine the surface current rather than relying on a current atlas, will create a branch in the scenario tree at the time where the drifters begin to be used. Editing the weather information with the SeaWeather editor will also create a branch in the scenarios which have used the data in their respective calculations. The user can indicate a specific time along a drift either by typing the time or by pointing at a point along the drift track plotted on the screen.

Defining a Search Area

Once a number of scenarios have been defined for an active search, the search planner must choose a probable area to be searched. The user selects a set of scenarios he considers most probable from the scenario tree and a time in the future at which the search will be conducted. This defines a set of leaves in the tree, each with their respective areas of probability. These areas must be divided among the search units available in such a way as to maximize the probability of detection and minimize the cost in manpower and time.

The search planner "drags" an icon representing a particular search unit from his store of available resources to each area. Dropping the icon at the "commence search point" (CSP), the search planner then defines the search pattern to be followed by the unit. The characteristics of each unit include their speed of travel, effective search sweep width, and search endurance. The computer determines the search capability for each unit (which will also depend on how far the unit must travel to the assigned CSP), and graphically displays the coverage achieved by shading the area of probability to reflect the reduced likelihood of the target being found there.

Once a suitable search coverage has been achieved and the search units required identified, the search planner will dispatch the units to their respective CSPs. As searching progresses, feedback from the search units determines the actual paths taken and the coverage realized. The areas searched will modify the areas of probability for any scenario which places the target in the area during the search. This modification extends to scenarios which are defined later so

that the search planner can readily see situations made unlikely because of searches already conducted.

Conclusions

CANSARP V3 is designed to run on a Sun Sparcstation 1/GX, which features 12.5 MIPS and 1.4 MFLOPS, a 16 inch colour display, and a mouse pointing device. With this computing power driving the user interface, a very intuitive graphical interface is within reach. The speed of computing also allows more accurate oceanographic modelling of current seastate in real-time. With SunWindows software standard on the Sparcstation, the mechanics of a powerful, dynamic user interface are readily provided.

References

- [1] C.T. Niwinski and D.O. Hodgins, "Improved Canadian Search and Rescue Planning (CANSARP) Methods for B.C. Coastal Waters." Unpublished report by Seaconsult Marine Research Ltd. for Canadian Coast Guard, 1985.
- [2] S.S. Salvador and J.A. Stronach, "DRIFTALC: A Computerized Search and Rescue System for the Straits of Georgia and Juan de Fuca." Unpublished report by Seaconsult Marine Research Ltd. for Canadian Coast Guard.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

WAVEFORM DESIGN FOR
A HIGH FREQUENCY OCEAN SURVEILLANCE RADAR

R.H. Khan
Centre for Cold Ocean Resources Engineering
Memorial University of Newfoundland
St. John's, Newfoundland, A1B 3X5, CANADA

and

D.K. Mitchell
Northern Radar Systems Limited
P.O. Box 8323, Stn. A, Bartlett Building
St. John's, Newfoundland, A1B 3N7, CANADA

ABSTRACT

The authors are presently engaged in the design of an experimental long range, high frequency (HF) over-the-horizon (OTH) ground wave (GW) radar system. Motivated by the need to maximize the transmitter duty cycle in order to detect long range targets, the radar design will be based on a Frequency Modulated Interrupted Continuous Wave (FMICW) type waveform. This paper analyses the Ambiguity Function (AF) of FMICW waveforms and proposes a practical interrupt sequence.

INTRODUCTION

Northern Radar Systems Ltd. (NRSL) is involved in the construction of a shore-based, long range, HF radar facility for iceberg detection, ship tracking and ocean environmental measurements. At HF frequencies, electromagnetic waves are capable of OTH propagation in either the GW or sky wave mode. The NRSL prototype system relies on the GW mode of propagation, and has a design capability for OTH detection and tracking of targets up to 400 Km from shore with a range resolution of 400 metres [1]. The main objective of this contribution is to develop techniques for the radar waveform analysis and design to achieve the desired specifications for the prototype radar.

In a conventional single frequency pulsed radar system, the desired range resolution determines the pulse length. The maximum range is determined by the repetition rate of the pulses and the hardware limit on the maximum achievable peak power. Pulse compression techniques [2] may be used to realize high energy pulses with low peak powers. Whereas pulsed radars have low duty cycles, Frequency Modulated Continuous Wave (FMCW) radars transmit a continuous waveform. The range measurement of FMCW radars is measured as the difference in instantaneous frequency between the transmitted and received waveforms. The received waveform from a stationary target, with a two way travel time of (t_r) sec, differs in instantaneous frequency from the received waveform by $(f_r = \alpha t_r)$, where α is the sweep rate. FMCW radars have a range resolution competitive with pulse radars when the same signal bandwidth is used.

The FMCW waveform, with a 100% duty cycle, is ideal for bistatic or low power monostatic operation. For the high power monostatic application of NRSL, there are two problems with the FMCW approach. First, the necessary isolation between the transmitter and receiver for continuous operation at HF frequencies is very difficult to achieve. Secondly, due to the propagation losses, the signal levels decrease rapidly with distance and the dynamic range of the signals exceeds that of any available receiver hardware.

To overcome both of the above mentioned problems, an FMICW waveform [3] has been chosen for the prototype radar. This is an FMCW waveform gated on and off with a well defined sequence. The introduction of these interrupt sequences complicates the analysis of the radar performance in terms of the well known AF. With an FMICW radar, signals are only received when the transmitter is off. Thus, only a fraction of the reflected waveform is actually received.

The form of the AF for FMICW radars has been discussed in the literature [4]. The inverse problem of determining the parameters of an FMICW waveform to realize a desired AF is more difficult. Some simplification is possible for a GW-HF surveillance radar where many targets of interest are slow moving and the zero-Doppler AF is sufficient to evaluate the performance of the waveform. This zero-Doppler AF may be easily obtained from the general form of the AF and offers an insight into the effects of different interrupt sequences on the FMICW waveform.

Both random and deterministic gating sequences are considered. The random sequence is seen to result in an AF with a significant noise floor which may obscure weaker targets, whereas the deterministic sequence causes ambiguity pop-ups. It is shown here that these ambiguity pop-ups may be successfully controlled, resulting in an AF with a very low noise floor; also the instantaneous signal dynamic range is compatible with available receivers.

AMBIGUITY FUNCTION FOR FMICW WAVEFORM

Modern radar technology relies on the AF to determine the best waveform for a particular application [2]. For a radar waveform $u(t)$, the AF is generally defined in terms of the response function:

$$AF = \chi(r, \phi) \chi^*(r, \phi)$$

$$\text{where } \chi(r, \phi) = \int_{-\infty}^{\infty} u(t+r/2) u^*(t-r/2) e^{-j2\pi\phi t} dt \quad (1)$$

The response function has two parameters, time delay (r) and Doppler (ϕ) , and is a measure of the correlation of the waveform $u(t)$ with a delayed and/or Doppler shifted version of itself. This is used to evaluate the capability of a radar waveform to resolve targets close to each other in range and/or Doppler.

It is important to note that the receivers must always be off whenever the transmitter is on, and that both the transmitter and receiver have finite switching times. Thus, the transmitter and receiver interrupt sequences are not identical. The transmitted waveform is not received in its entirety and the radar performance needs to be evaluated by the "cross response" function between the transmitted and received signals.

Evaluation of the Response Function Components

The derivation of the AF, presented here, evaluates the AF in terms of a number of component response functions which can be combined linearly. A single period of an FMICW waveform may be described as the time domain product of an ideal infinite FM sweep (chirp) waveform $g(t)$, infinite transmit and receive waveforms ($A_t(t)$, $A_r(t)$), and a gate function with a finite width (T_g).

The transmit and receive waveforms are respectively described as:

$$0 \leq A_t(t) \leq 1, -\infty < t < \infty \quad (2)$$

$$0 \leq A_r(t) \leq 1, -\infty < t < \infty \quad (3)$$

$$A_g(t) = 0, \text{ if } A_t(t) \neq 0 \quad (4)$$

The chirp waveform may be expressed as

$$g(t) = K e^{-i\alpha t^2/2} \quad (5)$$

where: K is a scaling constant and α is the frequency sweep rate of the "chirp" waveform. The response function of $g(t)$ may be easily evaluated using equation (1):

$$X_g(\tau, \phi) = |K|^2 2\pi \delta(\phi - \alpha\tau) \quad (6)$$

The response function of the product of two time-domain signals may be evaluated by convolving their individual response functions along the Doppler variable. The transmitted waveform $A_t(t)$ may be expressed as a gated version of an ideal "chirp" waveform. The received waveform $A_r(t)$ may be expressed as the transmitted waveform delayed by the two way travel time (t_0), and further gated by the received gating sequence.

$$f_t(t) = A_t(t) g(t) \quad (7)$$

$$f_r(t) = A_r(t) A_t(t-t_0) g(t-t_0) \quad (8)$$

$$f_g(t+t_0) = A_g(t+t_0) A_t(t) g(t) \quad (9)$$

Examination of equations (7) and (9) show that the response function for an FMICW waveform may be evaluated by:

- Evaluating the response function for $g(t)$.
- Evaluating the "cross response" function for $A_t(t)$ and $A_r(t+t_0) A_t(t)$.
- Convolving the two response functions along the Doppler variable.
- Convolving the response function of the infinite sequences with the impulse response of the finite bandwidth.

The response function of the infinite "chirp" waveform is given by equation (6). The "cross response" function for the transmitter and receiver gating sequences can be evaluated using:

$$X_r(\tau, \phi, t_0) = \int_{-\infty}^{\infty} A_r(t+\tau/2) A_g(t+t_0-\tau/2) A_t(t-\tau/2) e^{-j2\pi\phi t} dt \quad (10)$$

The response function of equation (10) may be convolved along the Doppler variable, with the response function of the ideal "chirp" waveform to give the response function of the FMICW waveform. This ideal response function may now be broadened to account for the finite bandwidth of the FM sweep.

Zero Doppler Ambiguity Function

In many applications of HF radars, the target velocities are very small and the zero-Doppler AF is sufficient to evaluate the performance of the waveform. Convolution of the response function of equation (10) with the response function of the ideal chirp of equation (6), at each value of the delay parameter, will shift the response function (10) in frequency. This frequency shift is a linear function of the delay parameter (τ). A little reflection will show that the response function needs to be evaluated for a set of delay-Doppler parameters such that the values will be mapped to the zero Doppler axis after convolution with the simple response function (6) of the "chirp" waveform. For this set, the Doppler frequency corresponding to any delay τ is simply $(-\alpha\tau)$. Thus, the zero-Doppler response function for an FMICW waveform may be computed using:

$$X_z(\tau, \alpha, t_0) = \int_{-\infty}^{\infty} A_r(t+\tau/2) A_g(t+t_0-\tau/2) A_t(t-\tau/2) e^{j2\pi\alpha\tau t} dt \quad (11)$$

From a computational point of view, equation (11) may be simplified by replacing t with $t-t_0$:

$$X_z(\tau, \alpha, t_0) = \int_{-\infty}^{\infty} A_r(t+\tau) A_g(t+t_0) A_t(t) e^{j2\pi\alpha\tau(t+t_0/2)} dt \quad (12)$$

Equation (12) may be interpreted as the correlation of two sequences $P(t, t_0)$ and $Q(t, \tau)$ given by:

$$\begin{aligned} P(t, t_0) &= A_r(t+t_0) A_t(t) \\ Q(t, \tau) &= A_t(t+\tau) e^{j2\pi\alpha\tau(t+\tau/2)} \end{aligned} \quad (13)$$

NUMERICAL EVALUATION OF AMBIGUITY FUNCTION

The following examples are evaluated using the specification for the NRSL prototype HF long range radar:

Sweep bandwidth, B	375 KHz
Sweep repetition period, T_r	0.524288 sec
Range resolution,	400 Metres
Unambiguous range,	409.6 Km

The examples are plotted as a function of range, where

$$\text{range} = c(t_0 + \tau)/2, \text{ c is speed of light} \quad (14)$$

as compared to the usual delay parameter of (τ), to reflect the variation of the AF with range.

Random gating sequences are conventionally chosen as a means to distribute the gating ambiguity. As an example, consider a normal distribution of the periods (T) of the gating sequences. Minimal periodicities are produced and a near white ambiguity noise floor is obtained as shown in fig. 1. However, this noise floor is only about 30 dB below the target level. This would seriously affect the radar performance for weak targets in the presence of stronger targets.

Random sequences are not the only means to distribute the ambiguity introduced by the gating. While deterministic sequences may produce pop-ups in the ambiguity function, they are at well defined locations, and the ambiguity noise floor between these pop-ups is negligible. A simple example of deterministic gating is shown in fig.2. Note that there is a substantial zero magnitude region, where there are missing ambiguity pop-ups.

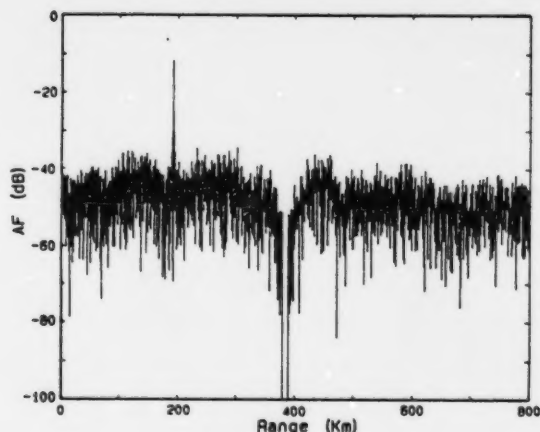


Figure 1: AF of FMICW with Random Interrupt Sequence, $t_0 = 196$ km.

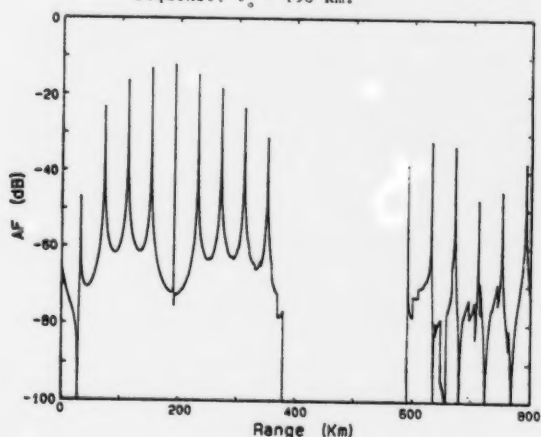


Figure 2: AF of FMICW with Deterministic Interrupt Sequence, $t_0 = T/2 = 196$ km.

FMICW WITH DETERMINISTIC INTERRUPT SEQUENCE

It remains to define analytically a deterministic sequence with no ambiguity pop-ups within a desired unambiguous range. Equation (10) shows that the "cross response" function for the gating sequence is evaluated by a one-dimensional Fourier transform of a two-dimensional function. Fig. 3a is an example of simple deterministic transmit and receive gating waveforms. The two-dimensional function of equation (10) is plotted in fig. 3b for these gating sequences. This is referred to here as a "diamond plot". The "cross response" function in fig. 3c is obtained as indicated in equation (10), by the Fourier transform of fig. 3b along the 't' variable. The function in fig. 3b is strictly periodic, with period (T) and of infinite length along the 't' variable. Thus, the "cross response" function has a discrete spectrum and can be easily evaluated knowing the Fourier transform of a periodic gate. The lines in fig. 3c represent the locations of the discrete spectral lines.

As described earlier, the zero Doppler AF is obtained by convolving the "cross response" function with the response function of the FM sweep. This is indicated by the zero Doppler mapping line on fig. 3c, which selects the values of the "cross response"

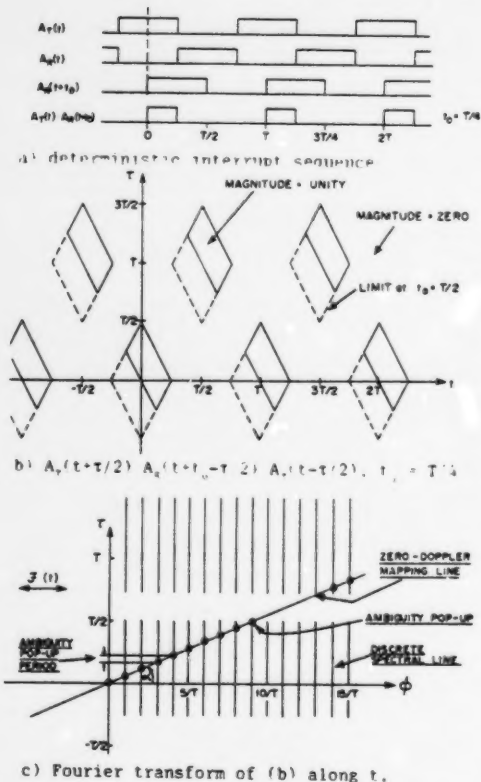


Figure 3: Response function of interrupt sequence

function which will be mapped to zero Doppler after convolution with the response function of the FM sweep. The ambiguity pop-ups will occur at the intersections of the zero Doppler mapping line with the discrete spectral lines of the "cross response" function. It can be easily shown that the ambiguity pop-up period is independent of target distance and is given by

$$T_a = 1 / (\alpha T). \quad (15)$$

An interesting effect shown in fig. 3c, is the zero magnitude region produced when the target distance (t_0) is less than one half the gate period ($T/2$). This effect is produced when the height of the diamonds is less than the repetition period of the gating sequence. This is the case where the product of the transmit gate and delayed receive gate has a duty cycle less than one half, such as when the target distance is less than one half the gate period, as in fig. 3a. Also, this is the case where the transmit on time is less than the receiver on time.

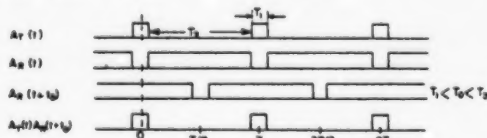
Consider the example of fig. 4a. For these gating sequences the receiver on time is significantly longer than the transmit on time. The diamond plot shown in fig. 4b is somewhat modified if the target distance is less than the transmit on time. However, this does not affect the repetition rate of the diamonds along the 't' axis and the zero magnitude response function in the range ($T_1 < t_0 < T_2$) is unchanged. Fig. 4b shows that a large unambiguous range may be achieved by:

- Ensuring that the transmit on time (T_1) is less than $1/(\alpha T)$.

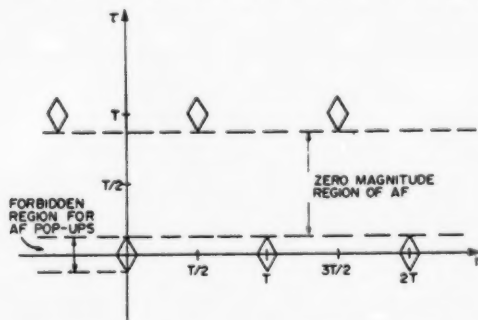
- b) Ensuring that the transmit off time (T_2) is greater than the unambiguous range.

For the NRSLS specifications listed earlier, employing equation (15) and the above two conditions, the transmit on time may be chosen as 67 Km and the transmit off time as 400 Km. These, of course, are for ideal conditions. Limited receiver bandwidth, transmitter and receiver switching times will call for modified values. The results of a computer simulation is shown in fig. 5. The extremely low value of the ambiguity noise floor will allow targets, well separated in signal power, to be detected easily.

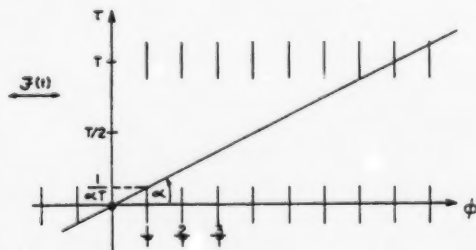
It has been shown in fig. 5 that a deterministic sequence may be designed with no ambiguity pop-ups over a desired unambiguous range. The cost is a reduced signal to noise ratio (SNR) due to a reduced transmitter duty cycle. For the example considered above, the signal power is reduced 8.4 dB from ideal FMCW. For the normal distribution random sequence considered earlier, the SNR would be reduced 3 dB from FMCW. A pulsed radar, to provide the required resolution of 400 metres with a maximum unambiguous range of 400 Km, would have a signal power reduced 30 dB from ideal FMCW.



a) deterministic gating sequence with low duty cycle



b) diamond plot with significant zero magnitude region



c) response function (note unambiguous region $0 < t < T_2$)

Figure 4: Design of Unambiguous Response Function

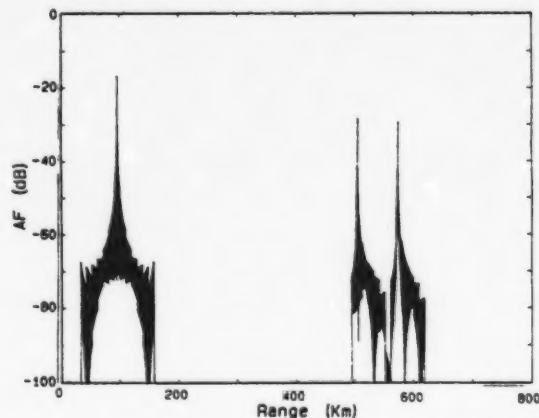


Figure 5: AF of Deterministic Sequence with Unambiguous Region

CONCLUSIONS

Analysis and design procedures have been developed for FMCW waveforms to be used in a prototype long range ocean surveillance radar. Random interrupt sequences, as typically used in FMCW systems, result in a substantial AF noise floor. However, simulations have shown that FMCW may be employed for HF OTH radar with a deterministic interrupt sequence. The specified resolution and unambiguous range for the prototype radar can be achieved with 8.4 dB duty cycle loss compared to an ideal FMCW system. Also, the required dynamic range of the receivers is now limited to a range equal to the width of the transmitter pulse, as is the case for pulse radar. These FMCW waveforms will be implemented shortly in hardware and the performance results reported accordingly.

ACKNOWLEDGEMENTS

Part of this work was carried out under contract between NRSLS and C-CORE. One of the authors (RHK) would like to acknowledge the support from NSERC in the form of an operating grant.

REFERENCES

- [1] C-CORE/MUN & Nortek Engineering Ltd., "Proposed Northern Radar Systems Limited Ground Wave Radar," Contract Report to NRSLS, 1987, Proprietary.
- [2] C.E. Cook and M. Bernfeld, "Radar Signals: An Introduction to Theory and Applications," New York, Academic Press, 1967.
- [3] E.D.R., Shearman, G.D. Burrows and M.D. Moorhead, "An FMCW Ground-Wave Radar for Remote Sensing of Ocean Waves and Currents," IEE Conference Radar 1987.
- [4] A.W.V. Poole, "On the Use of Pseudorandom Codes for "Chirp" Radar," IEEE Trans. Ant. Prop., Vol. AP-27, pp. 480-485, 1987.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

THE POTENTIAL USE OF RADAR TECHNOLOGY

IN SHIP BASED SEARCH AND RESCUE OPERATIONS

B.R. Dawe
NORCOO Limited
Donovan's Industrial Park
P.O. Box 8833
St. John's, Newfoundland
A1B 3T2

M. Kelly
Search and Rescue, Canadian Coast Guard
Canada Building, Minto Place
344 Slater Street
Ottawa, Ontario
K1A 0N7

C.A. Versailles
Transportation Development Centre
Guy-Favreau Complexe, Suite 601
200 Rene Levesque Blvd.
Montreal, Quebec
H2Z 1X4

Abstract

A sensor which can provide effective 24-hour search capability in the marine environment virtually independent of weather or lighting conditions is extremely attractive to those responsible for locating people and equipment in the Canadian marine environment. Radar Technology has the potential for providing this capability. Radar search and rescue (SAR) target detection experiments have demonstrated that standard ships' navigation radars are of limited use against common SAR objects. The major limiting factors in detecting such targets are that at longer ranges they are physically obscured from radar view by ocean waves for significant portions of the time and standard sea clutter reduction techniques and the radar display itself provide insufficient discrimination against sea return in the near range. The need for powerful radar signal processing and radical changes to the radar system is clear. Techniques employed by the most sophisticated of the military radars are likely to be necessary. Experiments with SAR radar transponders demonstrate that they are extremely effective as SAR radar target enhancement devices. Passive enhancement will only be effective if a reflector can be deployed at a significant height above the ocean surface.

Introduction

Since 1984 Transport Canada has been actively supporting research in the area of the technology for detecting small objects at sea from both ships and aircraft (see, for example, [1] and [2]). The ultimate aim of such work has been to increase the safety of human lives at sea through improved effectiveness of search and rescue (SAR) operations. A sensor which can provide 24-hour search capability in the marine environment virtually independent of weather or lighting conditions is extremely attractive to those responsible for locating people and equipment in the marine environment. Radar technology appears to have the potential for providing this capability. NORCOO Limited, under contract to the Transportation Development Centre and the Canadian Coast Guard (CCG) has undertaken a series of experiments, which, aside from other investigations, involved the evaluation of marine radars and an airborne radar [3] [4] [5]. The results of these evaluations are examined here with a view towards the insight they provide for the development of improved shipboard radar systems. Airborne radar detection results are provided to demonstrate the benefits of improved system characteristics and signal processing.

The Data

The data presented here were collected using a standard methodology developed by the United States Coast Guard (USCG) for such experiments. A detailed description of the data collection, reduction, and analysis methodology is provided by Dawe et al [6].

To summarize, a number of moored targets were distributed over an offshore experiment area. Simulated searches were undertaken during which a radar operator logged radar detections as the ship or aircraft proceeded through the array in a search pattern. Knowing the ship or aircraft and target positions at all times, it was possible to identify valid detections and opportunities. The lateral range or closest point of approach (CPA) was calculated for the opportunities and detections. Subsequently, plots of the probability of detection versus CPA or lateral range were developed.

Figures 1 and 2 show example plots for two ship's radars (Sperry 340 and 4016 radars) used operationally on the Canadian Coast Guard Ship (CCGS) "Sir William Alexander" and a Raytheon Pathfinder ST radar temporarily installed to evaluate its capabilities. The radar operators were CCG ship's officers. The reflectors were a Luneberg lens type having a nominal radar cross section of 2 square metres (m^2). When mounted inside the life raft they were believed to simulate reasonably well 4 to 6 people in a life raft. The centre of the reflector was about 0.5 m above the ocean surface. When mounted at canopy height it was about 1 m above the ocean surface. The antennas for the Sperry 4016 and 340 radars were situated at heights of 23.9 and 20.6 m, respectively. The Raytheon radar had an antenna height of 17.9 m. All three radars were very similar in their basic characteristics.

Figure 3 shows a plot of probability of detection versus lateral range for a Litton APS-504(V)5 airborne search radar deployed on an Atlantic Airways Limited King Air. The aircraft was flown at 150 m. The mode of operation was set at optimum for detection of this type of target, that is, frequency agility, scan-to-scan integration, constant false alarm rate processing, shortest pulse length (effective compressed pulse length of 30 nanoseconds), 20 nautical mile (nmi) range scale, and the highest antenna rotation rate (30 revolutions per minute (r/min) in this configuration).

Discussion

Shipboard Radar Detection Performance

Figures 1 and 2 do not give a true picture of the detection performance because of the rectification to CPA. The problem with sea clutter close-in is partially masked. Figure 1, however, certainly shows a flattening in the near range in comparison with what one would expect in free space. Figure 2 shows a considerable drop. In reality no detections were made in the clutter zone on the Sperry radars.

The figures illustrate several phenomena. The first is the clutter effect already mentioned. The second is the short range at which the probability of detection falls to zero. Finally, the overall

probability of detection is extremely low at any range despite the significant cross section. Through repeated observations of the plan position indicator (PPI) radar display, and visual observations of the target, these effects are easily explained. The significant wave height during most of this data collection period ranged from about 1 to 2 m. At longer ranges these targets were physically hidden from the radar by ocean waves for a significant portion of the time. At these ranges the radar operator is unable to distinguish between a momentary detection or a noise spike on the display. For the reflector mounted at canopy height this effect was substantially reduced, as is evident by comparing Figures 1 and 2. As the range to the target decreases the period of time it is visible to the radar increases, consequently the probability of detection increases. However, as the target enters the sea clutter zone it is immediately rendered virtually undetectable because it is indistinguishable from the clutter. By this time the target is continuously visible to the radar and probably presents a stronger target than the surrounding clutter; however, the very limited dynamic range of the display does not allow the operator to distinguish the target from clutter by observing signal intensity (grey level information). It would also appear that standard sea clutter reduction techniques in use on current marine radars tend to eliminate the target almost as effectively as they do the clutter. In the real situation then the target is usually detected for a very short range interval which lies between the outer limit of the sea clutter and the range at which the target is detected often enough to be recognized by the operator.

The Raytheon Pathfinder utilized a raster bright display and a basic m-out-of-n detection employed on a pulse-to-pulse and scan-to-scan (3 scans) basis. While not obvious from the plots, a few targets were detected in the sea clutter zone and targets detected outside the zone could occasionally be tracked through the clutter zone. However, for the vast majority of the cases the situation as described above applied.

Airborne Radar Detection Performance

By comparing Figure 3 with Figures 1 and 2 the advantages of the sophisticated signal processing and increased antenna height are clearly evident. It is seen, however, that the overall performance in the trials reported here is still far from ideal. There is clearly a problem in the close in sea clutter zone. The problem was evident during the data collection period when observing the radar display. The physical visibility of the target should have been much less of a problem due to the height of the radar antenna. Also, during the data collection period a large number of false detection were logged by the airborne search radar operator. Indeed, only 15% of the targets detected were valid targets.

A USCG airborne search radar, a Texas Instruments AN/APS-127 system, was evaluated during the same experiment as the Litton APS-504(V)5 radar. It is a less sophisticated radar and does not incorporate pulse compression or scan-to-scan integration. The only significant system parameters it possesses are frequency agility and a 120r/min antenna rotation rate. The system is flown on a USCG HU-25 aircraft (Falcon jet) and in this particular trial the altitude of operation was about 760 metres. No valid detections were logged during the evaluation.

Immediate Implications of Results

The ultimate goal of this research work was to provide the CCG with radar performance information for use in real search operations. From the data collected a sweep width value is calculated. Sweep width is the maximum spacing (using a coverage factor of 1.0) between parallel search tracks during an actual search operation. For the marine radars a sweep width value of about 0.5 nmi was derived for life raft targets. This is an extremely small sweep width and one can appreciate that, based on this information, current ship's marine radars are not very effective search tools in the conditions encountered in these trials when searching for four- and six-man canopied life rafts. The airborne radar detection performance resulted in a sweep width of 5.25 nmi and is deemed potentially usable as a search tool for this type of target. The potential for a large number of false detections is cause for concern as frequent diversions from the search track to investigate each detection could substantially increase the time required to effectively cover a search area.

Considerations for Future Radar System Designs

The problem of target visibility from the point of view of shipboard radar design is difficult to overcome. For example, one proposed solution might be to mount the antenna as high as possible. This does of course increase the range at which sea clutter will become the limiting problem and this may override the advantage of the increased height. In addition, problems due to being above surface ducts may occur. Clearly there are no simple answers. It must first be recognized that marine radars are designed for navigation purposes and not for weak target detection. In order to attain reasonable performance for this type of target a combination of redesigning the radar system itself along with the addition of signal processing, which includes powerful clutter reduction techniques, is required.

While the basic signal processing on the Raytheon radar did not provide any significant reliable improvement in detection in the clutter zone the airborne radar results indicate that signal processing of a similar kind but implemented in a much more powerful manner could be of substantial benefit. Considerable insight into what might be utilized can be attained by investigation of military oriented radar, particularly, airborne search radars such as the APS-504(V)5. It is anticipated that the employment of signal processing techniques such as scan-to-scan integration, pulse-to-pulse integration, constant false alarm rate processing, and the utilization of system designs incorporating such attributes as pulse compression, high speed scanning, and frequency agility, will substantially improve the effectiveness of ship based radar search operations.

Williams [7] in a discussion of the detection of growlers (glacial ice) and other very small targets on ship's radars, suggests the need for two radar antennas making use of height diversity to allow detection out to the maximum range (over intervening waves) and to ensure the ability to radiate within any surface duct that may be present. Williams felt that height diversity was a far more cost-effective strategy than a single radar no matter how sophisticated it is. He also recommended the use of high speed scanning, powerful anti-sea

clutter processing, and the use of a median detector.

In the work reported here the improved performance of the airborne radar was due to the combination of both altitude and signal processing. Even with the clearly advanced design, this radar's performance cannot be considered solid against this type of target. Had the system been able to operate using an antenna rotation speed of 120 r/min its performance may have been significantly improved. Certainly the difference in the performance of the two airborne radars demonstrates the benefit of the signal processing employed by the APS-504(V)5 radar.

The radar display employed will play a critical role in the performance any specially designed radar system. Williams [8] [9] has demonstrated and discussed the impact of various types of displays. From this work the use of digital multilevel displays seems to be the best option.

Target Enhancement

Dave et al [4] reported the attributes of passive reflector enhancement and active radar transponder enhancement of SAR targets. In these trials the transponder was clearly shown to be the superior enhancement technique, giving a sweep width of almost 9 nmi for marine radar. In recent trials [5] it was found that an airborne search radar could only detect the transponder when it was operated in a non-optimum mode (longer pulse length and no frequency agility). The implications of this have to be considered when engaged in a true search where a transponder may or may not be in use. The cost to the user of the transponder and perceived potential liabilities of manufacturers may tend to limit widespread use of transponders at the present time.

Passive radar enhancement may be effective if a reflector of sufficient radar cross section (say $2m^2$) can be deployed at such a height as to overcome wave obscuration. The data presented here clearly demonstrates the value of increased height. Sea clutter still presents a problem in the near range. Again cost to the user may be a factor.

The utilization of passive, polarization rotating, reflectors may be of considerable benefit. Such devices have been developed at the Communications Research Laboratory, located at McMaster University, for navigation applications. A dual polarized radar system would be required for SAR purposes. Such a radar would, for example, transmit horizontally and receive the vertically polarized return from the reflector. Very little return would be received from the sea clutter (only the depolarized component). This technique might improve longer range detection as well.

Conclusions

Radar technology certainly holds good potential for SAR operations in Canada's offshore marine environment. It is clear, however, that considerable research and development is necessary to develop an optimum shipboard radar. Such systems may employ very sophisticated designs similar to those currently employed by the military. These systems will be expensive.

Target enhancement using active radar transponders is a very attractive option. Passive enhancement could be an effective alternative as well. The willingness of the user to pay the

additional costs of these devices has to be considered. Unless all users are prepared to include active and/or passive radar enhancements, search vessels and aircraft utilizing radars must still be equipped with systems designed to reliably detect unenhanced targets.

Acknowledgements

This paper is based on data collected and observations made under work funded by the CCG and the Transportation Development Centre. The authors wish to thank the staff of the CCG, Maritime Region and the officers and crews of the CCGS "Sir William Alexander" and "Alert" for their support and cooperation. We also thank the USCG for its support and participation.

References

- [1] B.R. Dawe, D.J. Finlayson, and R.A. Stacey, "Evaluation of the Technology for Detecting Small Objects at Sea: Surface Sensor Platform," Transport Canada Report No. TP6818E, 1985.
- [2] R.D. Worsfold, "Evaluation of the Technology for Detecting Small Objects at Sea: Airborne Sensor Platforms", Transport Canada Report No. TP7359E, 1985.
- [3] B.R. Dawe, D. Finlayson, and D. Bryant, "Search and Rescue Detection Experiment Placentia, Nfld.," Transport Canada Report No. TP8028(E), September 1985.
- [4] B.R. Dawe, D. Finlayson and D. Bryant, "Search and Rescue Detection Experiment - Canso Bank, 1987," Transport Canada Pub. No. TP9475E, 1988.
- [5] B.R. Dawe and D. Finlayson, "Search and Rescue Detection Experiment - Canso Bank, 1988," in press.
- [6] B.R. Dawe, D. Finlayson, D. Bryant, "Results of a Canadian Shipborne Radar Search and Rescue Detection Experiment," in Proceedings of Oceans '88, 1988, pp. 1427-1432.
- [7] P.D.L. Williams, "The Detection of Ice at Sea by Radar," The Radio and Electronic Engineer, Vol. 49, no. 6, pp. 275-287, June 1979.
- [8] P.D.L. Williams, "The Impact of New Display Technology on the Detection of Small Targets at Sea by Radar," Electronic Circuits and Systems Vol. 3, No. 6, pp. 241-246, November 1979.
- [9] P.D.L. Williams, "Observations on the Further Optimization of Radar Signal Processing for the Display and Detection of Targets in Sea Clutter," International Journal of Remote Sensing, Vol. 5, No. 6, pp. 489-496, 1984.

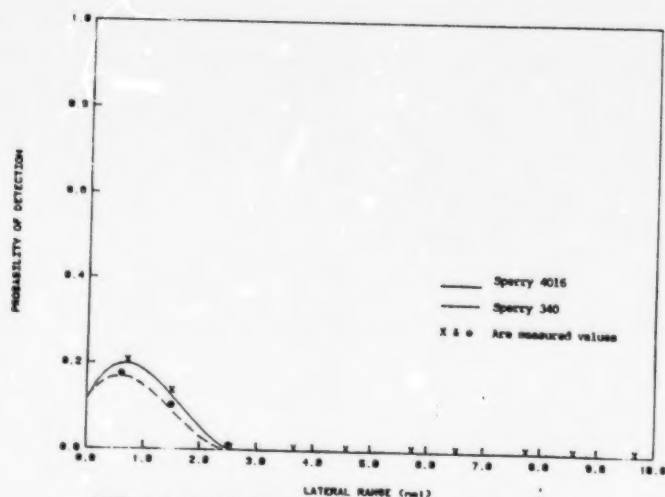


FIGURE 1 Derived Radar probability of Detection versus Lateral Range against 4 and 6-man Life Rafts with 2 square metre Reflectors Mounted Inside [6].

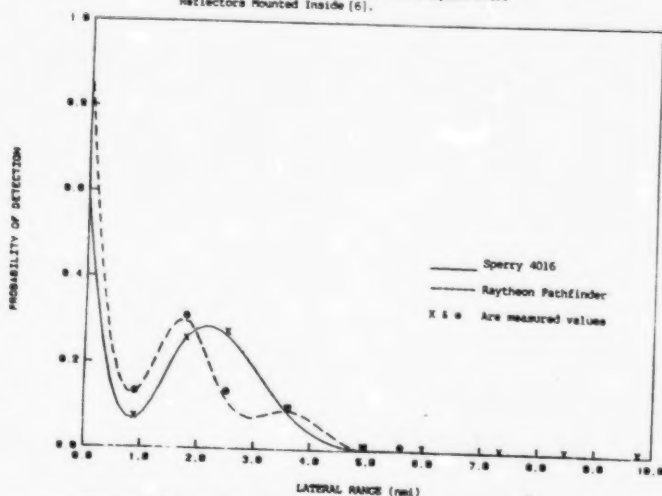


FIGURE 2 Derived Radar Probability of Detection versus Lateral Range against 2 square metre Reflectors Mounted at 4-man Life Raft Canopy Height [6].

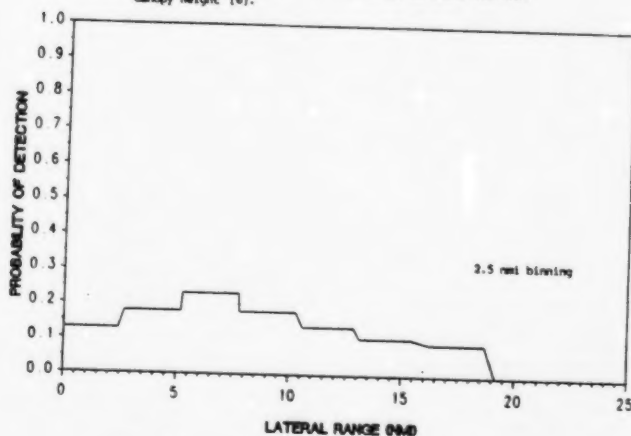


FIGURE 3 Probability of Detection versus Lateral Range for the Litton APS-504(V)5 Airborne Search Radar against 4 and 6-man life rafts with 2 square metre reflectors mounted inside [5].

**ADVANCED SYSTEMS FOR AIR SPACE
MANAGEMENT/
SYSTÈMES AVANCÉS DE GESTION DE L'ESPACE
AÉRIEN**

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

A TAXI AND RAMP MANAGEMENT AND
CONTROL SYSTEM (TARMAC)

U. Völckers

German Aerospace Research Establishment (DLR)
Institute for Flight Guidance
Flughafen, P.O.Box 3267
D-3300 Braunschweig, Germany

The air traffic control system is in need of improved airport surface traffic control system. It has to meet a variety of requirements of the airport operators, the airlines, pilots and ATC controllers.

Most significantly any new system has to be an integral part of the overall future Air Traffic Management System. All aspects of arrival and departure flow control as well as airport facilities management have to be considered, in order to allow safe and efficient ground operations of aircraft. A new operational concept is under development at DLR, which does not only necessitate new solutions in hardware or software functions but most important a very careful new design of the ground controller's working environment.

1. Introduction

In the past decade ground-based and airborne systems and the necessary procedures have been developed and introduced, which allow all-weather approaches and landings. More and more aircraft and airports are equipped for Cat IIIa/b operations. However, while low-visibility landings have become a normal procedure, problems arise, once the aircraft are on the ground. In low-visibility neither the Tower Controllers nor the pilots have precise information about the position of the aircraft nor can precise guidance information be provided.

This results in reduced operational capacity and possible safety risks. As long as only a few modern aircraft were CAT III equipped, the capacity issue was not crucial, as there were only a few aircraft able to land in low visibility. With the introduction of the new generation of transport aircraft the issue of efficient and safe low visibility ground operations has become very important, if ground operations should not become the capacity bottleneck in ATC.

But even in good weather with up to 70, 120 or even 200 operations per hour at major airports, there is increasing need for efficient safe and expeditious ground operations.

The development of a Ground Movement Control System therefore has become one of the key issues in the development of the future Air Traffic Control Systems.

The development effort at DLR is not being conducted to find a solution based upon the application of a specific technology or existing system. On the contrary, a new Ground Movement Control system has to be of modular design in order to assure adaptability to different airport and traffic demands and compatibility with other technical developments.

2. What is Airport Ground Movement Control?

2.1 Definition

The Airport Ground Movement Control System is defined as that system (equipment, procedures, people) which is concerned with the movement of:

- o Arriving aircraft through the phases of landing and taxiing to the passenger/cargo terminals.
- o Departing aircraft through the phases of engine start-up, push-back from the parking position and taxiing to the departure runway and take-off.
- o Aircraft in transit between different sites at the airport (e.g. from passenger terminals to maintenance area or v.v.)
- o Service or emergency vehicles, (e.g. follow-me-cars, snow-ploughs and rescue vehicles) operating on the airport taxiways and/or runways.

2.2 Functions and Responsibility

In order to perform the safe and efficient movement of aircraft and vehicles on the airport movement surfaces, such as runways, taxiways and apron areas, several parties have to work together. The main parties involved in this task are:

- o Air Traffic Services
- o Apron Management Service
- o Pilots
- o Ground Vehicle Drivers.

Air Traffic Services

The Controllers in the airport TOWER are responsible for the control of aircraft from the landing runway to the parking position, on the apron and back again to the take-off runway, as well as other movement of aircraft and vehicles on the airport surface, such as from a maintenance area to an apron or from apron to apron. They have to establish routings for arriving and departing aircraft; issue clearances to aircraft, sequence movements on runways and taxiways to ensure an expeditious traffic flow; monitor the compliance with their instructions; detect and resolve potential conflicts and avoid collisions; control the movements of service or emergency vehicles. The tower controller's surveillance function in determining the position and identity of aircraft or vehicles is normally accomplished by visual observation, supplemented by position reports obtained by voice radio communication with the pilots or by other sensors that provide additional visual signals.

Apron Management Service

While Air Traffic Service controls the ground traffic on the movement areas (i.e. runways and taxiways) usually a special Apron Management Service is responsible for the control of aircraft and vehicles on the apron area to ensure safe separation. They allocate parking positions; issue start-up, push-back and taxi-clearances; provide marshaller service etc.

Pilots

Control instructions are sent from the controllers to the pilots via voice radio communication. Each pilot is responsible for the navigation of his aircraft,

according to the instructions given by the tower or apron controllers. Lights, signs and markings are installed on runways and taxiways and at intersections to aid the pilot along his route on the runways, taxiways and apron. The pilots navigation fully relies on visual clues. The pilot in the end is responsible of maintaining safe longitudinal and lateral separation while taxiing. In low-visibility conditions guidance usually is provided by a "follow-me"-vehicle that leads the aircraft.

Ground Vehicle Drivers

Ground Vehicle Drivers are also part of surface movement guidance and control. In addition to the follow-me cars already mentioned above, there are maintenance and emergency vehicles and tow-vehicles using the same movement areas as the aircraft. Drivers of those vehicles must comply with all regulations and instructions. They are responsible for exercising due care and attention, so as to avoid collision between their vehicles and aircraft, and between their vehicles and other vehicles.

2.3 System Elements

The Ground Movement Control System consists of four interrelated major functional elements:

- o Surveillance and Identification,
- o Communication,
- o Navigation and Guidance,
- o Control and Planning.

Surveillance and Identification

Surveillance is the process whereby Ground Control facilities acquire information on the position and identity of vehicles under their jurisdiction. The Ground Controller presently uses mainly visual observation, through the windows of their tower cab, as his primary means of surveillance. Video information may be provided to observe remote or shaded areas. Infrared sensors occasionally may provide additional information at night or in low visibility. Some airports are equipped with an Airport Surveillance Radar (ASDE). However, the performance of all those devices mentioned is very limited and is only used as additional information.

Communication

Communication between Controllers and aircraft or vehicles is carried out via voice radio link. According to the functional organisation of the Ground Control units and due to high frequency load several R/T-Channels have to be used with all negative implications. On many airports communication overload is a restricting factor which contributes to ground delays.

Communication between the different ground facilities is performed via telephone, intercom or data lines.

Navigation and Guidance

Guidance of individual aircraft is exercised by each pilot, who carries out the tasks of controlling aircraft velocity, route/track-keeping, turning, stopping, collision avoidance. Runway and taxiway lighting, signs and markings, stop-bars and clearance bars, are installed to aid the pilot. In low visibility follow-me-cars provide the necessary guidance to the pilot.

It is obvious that with

- o increasing number of traffic movements,
- o increasing operations under low-visibility conditions,
- o increasing complexity of runway/taxiway/apron topology

all those systems elements have to be improved.

A future Ground Movement Control System will need:

- o an improved surveillance and identification capability that will provide a precise picture of the ground traffic activity independent of visibility conditions,
- o an improved planning and control capability, that will provide controllers with computer assistance to establish a safe, efficient flow of ground movement, including automatic conflict alert for runway incursions and at certain intersections. This planning system has to be closely interconnected with other ATC functions;

- o more efficient methods for controller-pilot communication,

- o an improved, precise and unambiguous taxi-guidance information to pilots.

Control and Planning

Control activities, such as determination of taxiway routings, taxi sequencing or priority at intersections, issuance of instructions, monitoring etc. presently is performed by human controllers who have to consider various information from a variety of different data sources. This is a solely mental monitoring, planning and decision making process, with no computer assistance. The overall Ground Control is divided into several controller positions, who have to coordinate their plans and actions. Furthermore efficient coordination with Approach Control, Departure Control and Airport Operations Management is urgently required to conduct efficient and safe Ground Movement Control.

3. Requirements for a Future Ground Movement Control System

The idea of this paper is to present a general concept for a Future Ground Movement Control System. This general concept or framework should be comprehensive and consistent. However, it should not be confined to certain technical systems or solutions, but should leave the chance for different options. Thus, the following requirements have been set-up as guideline for the design of the System Concept:

3.1 General Requirements

The Ground Movement Control System should be:

- o of modular design, so as to be
- o adaptable to any airport structure
- o adaptable to any traffic volume
- o appropriate to serve under any weather conditions
- o allow up-grading in an evolutionary way and adapt to the changing state-of-the-art of systems and technology
- o flexibel enough to allow different technical solutions or combinations thereof

- o if ever possible be in compliance with ICAO regulatory documents
- o an integral element of the ATC/ATM system
- o reduce controller and pilot workload.

3.2 Operational Requirements

3.2.1 Surveillance

The system should provide:

- o complete information on the position, identity and progress of all aircraft, aircraft under tow, as well as control and emergency vehicles which operate on the movement area (runways, taxiway, aprons,
- o information on stillstand, movement and the direction of movement of the targets,
- o information on the presence of temporary obstacles and other hazards.

3.2.2 Control and Planning

The system should provide:

- o information on the operational status of the elements
- o digital data interfaces to Approach Control, Departure Control, Flow Control, Airport Operations Management, Airline Operations Management
- o plans for expeditious operations such as: optimal start-up clearances, optimal push-back sequences, optimal taxi-routes to be followed
- o separation monitoring/conflict alert
- o conflict resolutions
- o comprehensive data displays and easy to handle M/M-Interface (Interactive)
- o digital data output in order to allow direct transmissions of instructions, etc. to the aircraft or for direct control of ground-based guidance facilities.

3.2.3 Guidance and Navigation

The system should provide:

- o unambiguous information on the present position
- o guidance on the route to be followed (ground-based or aircraft-based),
- o hazard warnings; areas to be avoided,
- o warnings, recommendations on direction changes, turns, stops, speed adjustment, etc.
- o information to prevent collisions or separation violations with other aircraft, vehicles or obstacles,
- o information on clearance violations (safety net).

3.2.4 Communication

The system should reduce voice radio communication by

- o increased automated data exchange between ground-based control units
- o increased automated transmissions (voice/data link) between ground-stations and aircraft or vehicles.

4. Proposal for a Modular System Concept

According to the requirements a modular systems concept has been designed. Starting from the present system "Step 0" (Fig. 4) and its functions, a step-wise enhancement with modified or new elements is proposed. The final version "Step 5" (Fig. 5) covers all elements of the proposed integrated Ground Movement Planning System. Intentionally this concept is not confined to the application of any existing or expected technology or systems, but leaves the chance to apply different solutions or combinations.

A brief description of the systems concept shall be given.

Characteristic of the present system (Step 0) is, that the Ground Controllers have to use a variety of more or less independent data sources, each of them having its own man-machine-interface.

The functional elements can only be briefly described in this paper:

Surveillance - direct visual observation; if applicable additional visual aids or airport surface radar; pilot reports.

Identification - by mental correlation of various information and perception.

Data Collection and Coordination - by telephone with adjacent ATC-units; by monitoring other radio frequencies; use of flight progress strips, information on parking position management. Data collection and coordination require a significant share of the controllers' mental capacity.

Planning - Mental planning of all necessary actions with consideration of operational procedures, Airport Structure and Status Data and Weather Data.

Communication - Voice radio communication to aircraft and vehicles to-be-controlled.

Guidance - according to instructions transmitted to the pilots/drivers. Use of maps, signs, markings, lights, etc. as guidance signals.

Man-Machine-Interface - Variety of separate, independent displays, controls, communication channels, lists etc.

The modular concept envisages the following steps:

Step 1: Integration of computer-assisted coordination with Arrival and Departure Control. Integrated Man-Machine-Interface.

Step 2: Introduction of computer-based Ground Operations Planning System (optimal taxi routing).

Step 3: Introduction of automated identification and tracking function, facilitating continuous automatic situation monitoring and dynamically updated optimized taxi flow instructions, using a central knowledge base.

Step 4: Introduction of automated short term conflict detection, warning and resolution functions.

Step 5: Introduction of Guidance Data Processing function which either should automatically

provide ground-based guidance signals or automatically transmit guidance information to the aircraft via data-link.

5. Concluding remarks

A general concept of modular design for a Ground Movement Control System which can be adapted to cover all necessary functions has been set-up.

Yet, the detailed solutions still have to be worked out.

We at DLR are proceeding in the development of the Planning System Element for ground movement control.

We are working on a detailed concept for an integrated system with all necessary data management functions and interfaces to other ATC units. (Flow Control, Airspace Management, Airport Operations Management and Airline Operations Management).

We have started to develop new planning software that will assist the Ground Controllers to establish a safe, and expeditious traffic flow on the ground, with full consideration of Air Traffic Control, ATM, Airline requirements. At the same time we have begun to work out the new Man-Machine-Interface and operational procedures.

In order to do our task we are in the process to build-up a Ground Movement Simulation, to study and develop the operational procedures, the necessary displays and controls for convenient Man-Machine-Interaction.

This planning system, however, is only one element of the total system. New solutions for surveillance, Guidance and Communication have to be developed simultaneously on order to achieve the envisaged benefits.

Studies, development and evaluations of order systems and procedures will have to be conducted parallel to our work in order to specify technical and operational requirements and constraints. The experience gained will be useful in the international standardisation process that is expected to come.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

AUTOMATED PROBLEM RESOLUTION PROTOTYPE IN AUTOMATED
EN ROUTE AIR TRAFFIC CONTROL

C. Vanessa Fong
The MITRE Corporation
1820 Dolly Madison Blvd.
McLean, Virginia 22102
U. S. A.

Abstract

Automated En Route Air Traffic Control (AERA) is part of the Advanced Automation System software currently being developed by the U. S. Federal Aviation Administration (U.S. FAA) to modernize the Air Traffic Control system. The implementation of AERA involves three packages -- AERA 1, AERA 2, and AERA 3. Automated Problem Resolution (APR) is a function within AERA 2 that generates resolutions to conflicts detected by the Automated Problem Detection function, an AERA 1 capability.

In order to prepare APR algorithmic specifications for the Advanced Automation System, an APR prototype was developed in the FAA/MITRE AERA 2 Laboratory at The MITRE Corporation in McLean, Virginia. This prototype is a tool designed for evaluating rules and algorithms for resolving aircraft-aircraft and aircraft-airspace conflicts in both radar and nonradar environments. The evaluation is performed with FAA field operational personnel.

The APR Prototype is implemented on a Symbolics Lisp Machine using an expert system shell -- Knowledge Engineering Environment (KEE). It has three major functions: Conflict and Resolution Information Display, Problem Detection, and Resolution Generation. This paper discusses the overall characteristics of the APR Prototype and, specifically, the Conflict and Resolution Information Display and the Resolution Generation capabilities of the Prototype.

Introduction

The U.S. FAA is in the process of implementing the National Airspace System (NAS) Plan [1] which is a comprehensive plan for modernizing the national Air Traffic Control (ATC) system to meet the air traffic loads of the 1990s and beyond. A major component of the NAS Plan is the Advanced Automation System (AAS) which entails new hardware and software to be used. AAS includes a group of ATC support functions that are collectively referred to as Automated En Route ATC (AERA). When fully implemented, AERA will do the following [2]:

- Allow most aircraft operating under Instrument Flight Rules (IFR) to fly fuel efficient profiles
- Reduce the potential for operational errors
- Integrate en route metering with local and national flow control
- Increase the number of aircraft and volume of airspace that a control team can safely manage; i. e., increase sector capacity

The AERA capabilities are to be implemented incrementally, in three implementation packages: AERA 1, AERA 2, AERA 3. AERA 1 introduces capabilities within the AAS software to detect problems corresponding to an intended route of flight, to display problem information to a sector controller, and to provide additional aids to assist the controller with ATC control and coordination tasks. AERA 2 expands AERA 1 capabilities in several areas and introduces capabilities within the AAS software to automatically generate resolutions to problems (that are detected by automated problem detection logic), to display the resolution alternatives to the appropriate controller, and to provide additional

aids to further assist controllers with ATC control and coordination tasks. AERA 3 will introduce automated decision-making and control into the ATC process, and will exploit the capabilities of advanced avionics, e.g., flight management systems (FMSs) -- more fully than is possible with earlier ATC systems. AERA 1 capabilities are to be implemented by the AAS Acquisition Phase (AP) Contractor as part of the initial Area Control Computer Complex (ACCC) in the 1990s; AERA 2 capabilities are to be implemented as an early enhancement to the ACCC and the installation of the AERA 2 software in the first ATC site will take place in the 1999 time frame; and AERA 3 capabilities are to be implemented as a later enhancement to the ACCC [2-5].

For AERA, a problem is defined as any one of the following two situations: (1) a possible future violation of separation minima between two aircraft (hereafter referred to as an aircraft conflict), or an aircraft and a protected airspace volume (hereafter referred to as an airspace conflict); and (2) a possible future violation of a flow instruction [2]. AERA 1 provides the automatic problem detection capability for these two types of problems. AERA 2 provides the capabilities of: (1) automatic resolution generation for detected problems, (2) computer-generated data link clearances for controller approval, and (3) automated coordination between controllers [2].

Considerable progress has been made in the definition of the AERA 2 concept [3, 4]; however, the need still exists to further develop some of the AERA 2 capabilities for which some technical and operational risk still remains. The FAA/MITRE AERA 2 Laboratory facilities in McLean, Virginia currently are being used to support evaluations in two areas of the AERA 2 concept that require additional development and validation -- AERA 2 Computer-Human Interaction (CHI), and APR. The results of the CHI and APR evaluations will produce detailed specifications for the AAS contractor.

The CHI validation activities, conducted through the use of a CHI prototype [5], focus on the definition, validation and demonstration of controller-automation interfaces and interactions for AERA 2. Interfaces are lists, displays, and messages which are defined in the AERA System Level Specifications [3]. The CHI Prototype is used to study the interface information content, operational relationship between displays, as well as the formats and presentation style of the interfaces. The validation of the rules/algorithms (hereafter referred to as the APR Knowledge Base) that provide for the automatic generation of resolutions capabilities in AERA 2 is being performed through the use of the APR Prototype. The Prototype characteristics will be discussed later.

APR Overview

For each conflict detected, APR generates a set of resolutions for each aircraft involved in the conflict. Each resolution generated is checked against the environment by the problem detection capability (an AERA 1 capability) to determine whether it will cause other conflicts. APR tries to generate resolutions that are "conflict-free": free from conflicts with the aircraft that is involved in the original conflict, and free from conflicts that involve any other aircraft in the environment, i.e., secondary conflicts (also called "objections"). If APR cannot produce a completely "conflict-free" resolution for an aircraft, it will then propose a resolution with the "furthest" secondary conflict attached. In addition, flow instructions are also being considered in the APR resolution generation process. The

resolutions generated by APR are ranked based on operational considerations and then displayed. Resolutions with secondary conflicts attached are ranked in one group which is always placed at the bottom of the displayed list of resolutions. See [2,6,7] for specific APR resolution ranking criteria.

The APR Knowledge Base attempts to generate resolutions in three maneuver dimensions -- the lateral (lateral flight path change), the vertical (altitude change), and the longitudinal (speed change) dimensions for each aircraft involved in a conflict. The APR lateral dimension resolutions generation attempts both a "change path left" and a "change path right" resolution; the vertical dimension a "climb" and a "descent"; and the longitudinal dimension an "increase speed" and "decrease speed". Therefore, APR attempts to generate one resolution from each specific dimension and specific direction for each aircraft which is involved in the conflict.

APR Prototype

The APR Prototype was developed by MITRE and resides at the FAA/MITRE AERA 2 laboratory. It provides a means for evaluating the APR Knowledge Base's ability to adequately resolve aircraft conflicts, airspace conflicts and flow instruction violations [2, 6, 7]. It is designed to allow an evaluation by a team of experienced FAA operational field personnel, the Air Traffic AERA Concepts Team (ATACT), to address the technical feasibility of the APR Knowledge Base and the operational acceptability of its results. The APR Prototype is being developed in four phases (hereafter referred to as APR Prototype Builds) in order to facilitate knowledge acquisition from ATACT and to support a focused evaluation of different aspects of the APR Knowledge Base's resolution generation capabilities. The APR Prototype and its associated functional capabilities are the focus of the remainder of this paper.

APR Prototype Environment And Usage

The APR Prototype is being used to evaluate and validate the APR Knowledge Base which consists of definitions, rules, algorithms, and parameters. It requires an environment that satisfies the following criteria: ATC knowledge representation and encoding, knowledge acquisition facilities, knowledge base modification capability, and timely software development support. These criteria were satisfied by taking advantage of Artificial Intelligence/Knowledge Based System technology and using an expert system shell -- Knowledge Engineering Environment (KEE), and Lisp on Symbolics Lisp Machines.

The entire APR Prototype hardware consists of 7 networked Symbolics 3650 workstations. Each workstation includes one black and white and one color display. This hardware environment was selected and configured to facilitate evaluation by ATACT.

The APR Knowledge Base is evaluated using scenarios that are based on the ATC environment and flight data associated with four ATC centers (Minneapolis center, Washington center, Atlanta center, and Dallas-Fort Worth center). Conflict situations are "crafted" by defining flights with crossing paths or sharing same altitudes. Conflicts are detected by the Automated Problem Detection (APD) capability within the APR Prototype. For each conflict detected, the APR Prototype generates a set of resolutions for display.

The APR Knowledge Base is evaluated using various evaluation "runs." Each run is designed to focus the evaluation effort on a specific area of the APR Knowledge Base. Each "run" is a scenario which consists of ATC system information, environment information, flight information and aircraft performance information. In addition, each evaluation run contains at least one conflict. The APR Prototype provides facilities to create conflict situations so that different conflict geometries can be "crafted" and evaluated.

APR Prototype Functional Characteristics

The APR Prototype has three major functions: Conflict and Resolution Information Display, APD [8,9], and Resolution Generation.

Conflict and Resolution Information Display: The current APR Prototype supports both black and white and color displays. The Prototype user interface was designed to facilitate ease of APR Knowledge Base evaluation and knowledge acquisition. The user interface was initially evaluated by ATACT to determine the types and format of data required for the evaluations. The following is a brief discussion of the data which are displayed by the Conflict and Resolution Information Display function.

The black and white display presents summaries of the conflict information and the set of resolutions generated by the Prototype. Each conflict is characterized by the aircraft/airspace involved in the conflict, conflict start and end times, and the amount of time remaining to conflict start time. Each resolution is characterized by its clearance, rating (defined as the time difference between the subject aircraft's planned arrival time and its new arrival time based on the resolution), and ranking. Also, the set of APR resolutions can be selectively retrieved by a specific dimension -- vertical, lateral, or longitudinal depending on whether a particular resolution involves an altitude change, a lateral flight path change, or a speed change maneuver. Furthermore, "failed resolutions" that were attempted by the APR resolution generation capability are also recorded and can be retrieved. This last capability enables a "trace" of the resolution generation process.

The color display presents textual and graphical information which supplements the conflict and resolution information presented on the black and white display. The textual data displayed on the color screen are: (1) flight plans of all the aircraft in the scenario; (2) airspace characteristics including name, altitude bounds, activation schedule, and adapted reroute, if any; (3) conflict characteristics; (4) resolution clearance; (5) aircraft altitude, speed, and heading data at a specified time; and (6) aircraft separation data including lateral distance and vertical altitude difference at a specified time. The graphical information shown on the color screen displays the scenario's ATC environment. It includes the boundaries for sectors and airspaces; routes and names of airways; names of fixes; aircraft identifications and flight paths, and current positions, altitudes, and speeds of aircraft. If there is a conflict in the environment and the conflict has been requested (by selecting the conflict item from the black and white screen), the conflict region is then labeled on the aircraft's flight path(s). Additionally, if a resolution has been requested (by selecting a resolution item on the black and white screen), the resolution's path is also shown.

APD: The APD function in the APR Prototype provides the capability for detecting aircraft conflicts, airspace conflicts, and flow instruction violations. Since APD is an AERA 1 function, it will not be discussed further.

Resolution Generation [2,6,7]: The APR Prototype Resolution Generation function has three subfunctions -- Lateral Dimension Resolution Generation, Vertical Dimension Resolution Generation, and Longitudinal Dimension Resolution Generation.

The APR Prototype Knowledge Base for the lateral dimension calls for different types of resolutions to be generated depending on the aircraft's navigation capabilities. If an aircraft is equipped with Very High Frequency Omnidirectional Radio (VOR), then six types of resolutions are attempted: Direct to Downstream VOR (Type 1), Direct from VOR (Type 2), Reroute through One VOR (Type 3), Reroute Through 2 VORs (Type 4), Heading Off Route followed by Direct to VOR (Type 5), and Two Vectors (Type 6). If an aircraft is equipped with Area Navigation (RNAV), i.e., a point-to-point (PTP) aircraft, APR then attempts the following: One-Leg, direct to a point downstream (Type 1),

Two-Leg (Type 2), and Three-Leg (Type 3, parallel route offset). All of these are "single dimension" resolution types which only involve maneuvering one aircraft in one dimension. If APR Lateral Dimension Resolution Generation fails to generate a "conflict-free" single resolution for one side (left or right side) for an aircraft, then APR attempts to generate "composite" maneuvers or "multiple" maneuvers. A "composite" maneuver involves maneuvering an aircraft in the lateral and vertical dimensions simultaneously. A "multiple" maneuver consists of two independent maneuvers for two aircraft, in one resolution. If APR Lateral Dimension Resolution Generation fails to generate a "conflict-free" composite or multiple resolution, the APR Lateral Dimension Resolution Generation will then propose a single dimension lateral resolution with the "furthest" secondary conflict attached. If no such resolution exists, APR proposes no lateral resolution for that aircraft for that side. Figure 1 illustrates the functional flow of the APR Lateral Dimension Resolution Generation process.

Depending on whether the aircraft will be climbing, level, or descending within a parameter distance beyond its position at conflict detection time, APR Vertical Dimension Resolution Generation attempts different resolution maneuver types: (1) For an aircraft that will be in level flight, APR attempts Initiate-Climb to transition the aircraft to a higher altitude, and Initiate-Descent to transition it to a lower altitude, (2) For an aircraft that plans to climb, APR attempts an Alter-Climb resolution type which alters the aircraft's original climb profile, and (3) For an aircraft that plans to descend, APR attempts an Alter-Descent resolution type which alters the aircraft's original descent profile. Similar to the Lateral Dimension Resolution Generation, the Vertical Dimension Resolution Generation also attempts "composite" maneuvers or "multiple" maneuvers if "single dimension" vertical resolution fails. The remaining Vertical Dimension Resolution Generation processes are the same as those for the lateral dimension. Figure 2 illustrates the functional flow of the APR Vertical Dimension Resolution Generation process.

The APR Longitudinal Resolution Generation is not attempted if a conflict is of the head-on geometry. Otherwise, APR Longitudinal Dimension Resolution Generation attempts both an Increase-Speed and a Decrease-Speed resolution. If either one

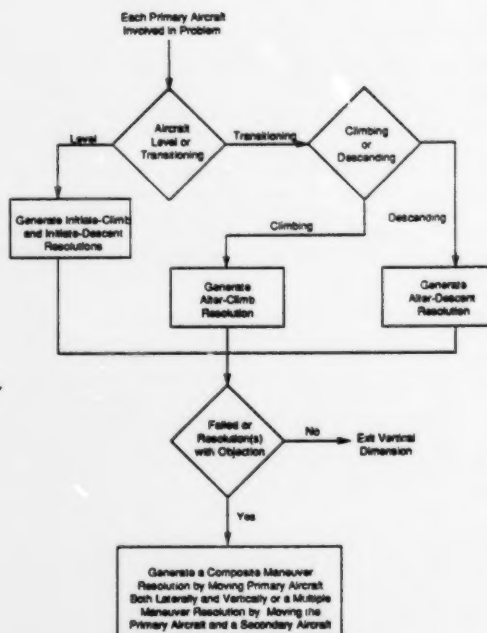


Figure 2
Vertical Dimension Resolution Generation Functional Flow

fails to generate a "conflict-free" resolution, APR Longitudinal Dimension Resolution Generation will further attempt a "multiple maneuver" by increasing the speed of one aircraft while decreasing the speed of the other. The remaining longitudinal resolution generation operations are the same as those in the lateral and vertical dimensions. Figure 3 illustrates the functional flow of the APR Longitudinal Dimension Resolution Generation process.

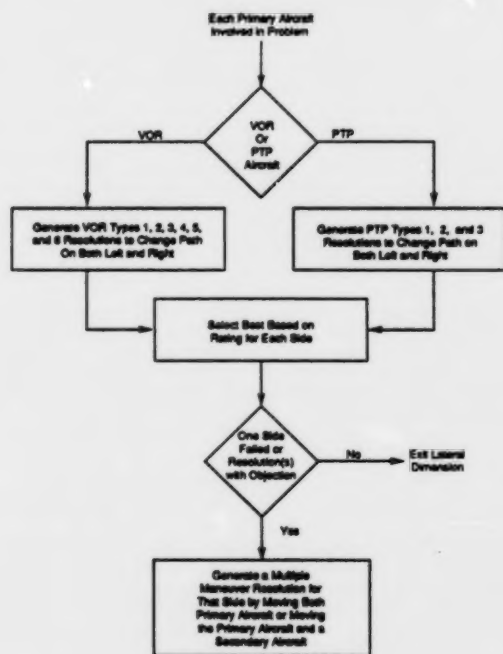


Figure 1
Lateral Dimension Resolution Generation Functional Flow

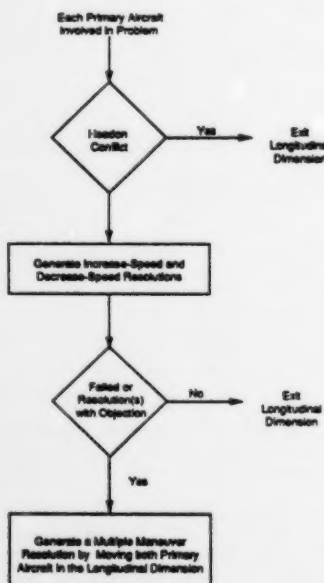


Figure 3
Longitudinal Dimension Resolution Generation Functional Flow

APR Prototype Status

The development and implementation of the APR Prototype involves four "Builds" -- Build 1, Build 2, Build 3 and Build 4. APR Prototype Build 1 was released and evaluated in April, 1988; Build 2 in September, 1988; Build 3 in March and May, 1989; and Build 4 will be completed in January, 1990. Each Build adds more APR functional capabilities and evaluation complexity to the APR Prototype. This "Build" approach allows incremental software development, and APR evaluations by ATACT at different levels for different issues. Additionally, new knowledge gained in previous APR Builds can be incorporated for validation in later Builds.

For the pre-Build 1 version of APR, strategy selection rules were used first to select applicable resolution strategies and associated aircraft to maneuver as part of those strategies. Once the applicable resolution strategies and their respective aircraft to maneuver were selected, maneuver type selection rules were applied to select applicable maneuver types from among those maneuver types associated with the selected strategies [6]. Rules were not used to define the specifics of individual maneuvers. Build 1 involved the evaluation of the strategy and maneuver type selection rules of the APR Knowledge Base for resolving aircraft and airspace conflicts in a simplified scenario environment - one pair of aircraft in a "windless" and radar environment [8]. The results from the Build 1 evaluation indicated that APR conflict resolution intelligence should be directed towards defining the specifics of individual maneuvers. Starting with Build 2, maneuvers are generated for all aircraft in all dimensions and all directions, eliminating the need for strategy selection rules and maneuver type selection rules. By using KEE and the Symbolics Lisp Machines, the prototype developers were able to rapidly implement the many changes in the APR Knowledge Base. Build 2 was evaluated in a more complex scenario environment than Build 1-- multi-aircraft and multi-airspaces. This Build was still limited to a "windless" and radar environment [9]. Build 3 included an APR Knowledge Base for the nonradar environment and was upgraded to handle wind layers [2]. The evaluation results of Build 3 confirmed the stability of the current APR Knowledge Base. APR Build 4 will include resolutions incorporating Flow Instructions (FI) and interfaces with the Traffic Management System.

References

- [1] FAA, 1986, National Airspace System Plan.
- [2] A. G. Chambliss, et al. (draft, scheduled for publication December, 1989), AERA 2 Automated Problem Resolution - Build 3 Knowledge Base, MTR-88W52, McLean, Virginia: The MITRE Corporation.
- [3] C. G. Ball, et al. (June, 1986), AERA System Level Specification, MTR-85W221, Revision 1, McLean, Virginia: The MITRE Corporation.
- [4] C. G. Ball, et al. (August, 1986), AERA 2 Operational Description, MTR-85W66, McLean, Virginia: The MITRE Corporation.
- [5] W. L. McCabe and J. F. McNelis (August, 1988), AERA 2 Program Plan, MTR-88W60, McLean, Virginia: The MITRE Corporation.
- [6] C. G. Ball and A. G. Chambliss (April, 1988), AERA 2 Automated Problem Resolution - Build 1 Knowledge Base, MTR-88W52, McLean, Virginia: The MITRE Corporation.
- [7] A. G. Chambliss, et al., (August, 1989), AERA 2 Automated Problem Resolution - Build 2 Knowledge Base, MTR-89W71, McLean, Virginia: The MITRE Corporation.
- [8] A. G. Chambliss (November, 1987), APR Functional Description, Build 1, WP-87W00364, McLean, Virginia: The MITRE Corporation.
- [9] A. G. Chambliss (October, 1988), APR Prototype Build 2 Functional Description, WP-87W00507, McLean, Virginia: The MITRE Corporation.

Acknowledgement

The author would like to acknowledge S. Pszczolkowski of the U. S. FAA and A. G. Chambliss, P. Y. Ryberg, and R. L. Brant of The MITRE Corporation who reviewed a draft version of this paper and provided many helpful comments.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

A UNIQUE APPROACH TO AIRCRAFT CONFLICT RESOLUTION USING
ARTIFICIAL INTELLIGENCE TECHNIQUES

I.G. Smith, C.A. Crabb, G.R. Ram, M.R. Attisha

CompEngServ Ltd.
Suite 600, 265 Carling Avenue
Ottawa, Ontario
K1S 2E1

Abstract

CompEngServ Ltd. has recently completed a project to develop and test a proof-of-concept, aircraft conflict resolution advisory system for non-radar airspace. This system, known as CORES (CONflict RESolution System), predicts losses of separation and presents ranked resolution alternatives.

Using a unique system architecture, CORES deals with the difficult problem of convergence and has been successfully tested using a wide range of scenarios.

This paper presents the logical architecture of the conflict resolution algorithm used in CORES, and discusses both the performance of the system and the direction of future developments.

Introduction

Air Traffic Control

Demands on air traffic control (ATC) facilities are ever increasing. As a result, controllers are in need of more sophisticated tracking and planning aids to provide the expeditious flow that the airlines demand, while still maintaining a high degree of safety. Artificial intelligence (AI) techniques are being investigated to provide some of these aids.

Due in part to its vast size, most of Canada's northern airspace is without civilian radar coverage. Currently, non-radar airspace is controlled using only aircraft flight plans and periodic position reports from aircraft. This requires a controller to keep a precise mental picture of the airspace at all times with the additional aid of a manual or electronic plotting board. Because of such factors as weather and distances between waypoints, there is ample room for error in estimating the positions of aircraft. Understandably, aircraft separation criteria are particularly strict in such an environment. Compounding all of these factors is Canada's geographic location as a convergence point for many overseas flights.

A conflict prediction and resolution advisory system would be of obvious benefit to controllers. However, before the design and production of such a system can occur, a suitable conflict resolution algorithm must be developed and proven to be practical. Careful integration with other ATC systems and extending the capabilities of peripheral functions is of no use if the conflict resolution algorithm itself produces poor or unreliable advice.

CORES Project

CompEngServ Ltd. has recently completed a project to develop a proof-of-concept aircraft conflict resolution system for non-radar airspace. This work was done under contract with the Transportation Development Centre from June 1988 to August 1989.

The design of this system, known as CORES (CONflict RESolution System), was based on a previous prototype conflict resolution system (described in [1], [2], and [3]) and proposed extensions to the prototype as reported in [4]. The previous prototype was incomplete in that only certain types of conflicts were handled and many of the low-level routines were missing (e.g., computing the details of specific maneuvers).

In addition to completing the missing portions of the previous prototype, the primary goal was to further develop and refine the extensions described in [4] in order to solve the problem of convergence. Convergence refers to the problem of an airspace containing some number of aircraft, where the result of maneuvering any one aircraft may have secondary effects (induced conflicts) which require more aircraft to be maneuvered. In this type of situation, it is not clear that the process of maneuvering aircraft can converge to a completely conflict free airspace due, in theory, to the exponential complexity of the problem. However, in practical terms the goal of always achieving a completely conflict free airspace is unnecessary. Since there is inherent uncertainty in any long term predictions of aircraft positions, it may be acceptable to induce an uncertain conflict (e.g., one which is several hours away) in order to resolve an immediate conflict.

In order to develop, test, and experiment with the conflict resolution algorithm, a stand-alone demonstration package was built which can graphically illustrate the use of the resolution algorithm (see [4] for a brief description of the demonstration package). The original prototype was built using an expert system builder that was deemed too limited to achieve the above goals. Instead, portions of the original prototype were reimplemented within CORES which was built using the 386 version of the KEE expert system builder from Intellicorp.

Resolution Algorithm Logical Architecture

The design of the conflict resolution algorithm was guided by three underlying principles:

- i) Minimal perturbation of the airspace: maneuver as few planes as possible and keep deviations from the preferred flight plans to a minimum.
- ii) ATC constraints: consider such things as the airspace structure, aircraft performance characteristics, and the number of times a plane has been previously maneuvered when deciding on resolution procedures.
- iii) Widening field of view: In a similar fashion to a controller, consider only the immediate factors involved with the predicted conflict first, followed by more global factors as necessary.

These three principles have been embodied in a multi-stage pipeline style process as shown in Figure 1. The first three stages (initial resolution, relax constraints, and convergence) deal with conflict situations of increasing complexity. Each of these stages can suggest ways of avoiding the predicted conflict.

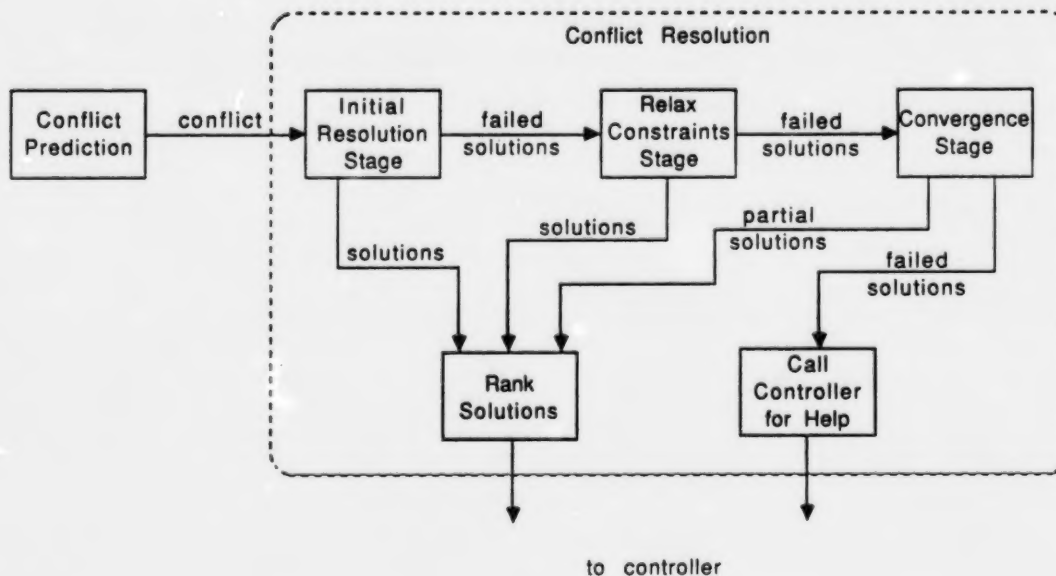


Figure 1: Main Stages of the Conflict Resolution Algorithm

Initial Resolution Stage

The initial resolution stage classifies the predicted conflict as one of six different basic types and selects strategies to test accordingly. The strategies tested during this stage are those that would typically resolve the majority of conflicts encountered. If a strategy can be executed such that no induced conflicts remain in the airspace then it is successful.

Relax Constraints Stage

Using those strategies that were tested by the initial stage but which failed, the relax constraints stage will attempt more extreme strategies. An example of a more extreme strategy would be routing a plane to a wrong-way altitude. Typically the minimal perturbation constraints are relaxed during this stage.

Convergence Stage

Finally, the convergence stage examines all those strategies which previously failed because they induced one or more further conflicts. In the convergence stage it is assumed that no strategies exist which produce a completely conflict-free airspace. The goal of this stage is to suggest partial solutions which result in a better overall airspace in terms of the criticality and difficulty of the induced conflicts. If the convergence stage fails to find any solutions which result in an improved airspace, the system will call the controller for help. Under these circumstances the system will provide the controller with all of the information known about why strategies have failed.

Ranking the Solutions

Solutions that are found by any of the three main stages are ranked and ordered according to suitability criteria and then presented to the controller. Some examples of the criteria used are, the expected deviation in flying time (as a result of the resolution procedure), the number of times an aircraft has been maneuvered previously in order

to resolve conflicts, and the ease of maneuvering the particular plane (weight class).

Conflict Prediction

An integral part of the conflict resolution system is the conflict prediction mechanism. The prediction mechanism provides the initial conflict information to the resolution algorithm and is also used to test each strategy proposed by the resolution algorithm. The computation of accurate resolution advice is dependent on accurate conflict prediction information.

Unlike other conflict prediction mechanisms, which typically test aircraft separation using constant or variable sized circles drawn around each aircraft, the conflict prediction mechanism in CORES uses a myriad of ATC procedural separation rules. This requires the prediction mechanism to fully analyze the conflict situation in order to recognize which rules to apply (e.g., crossing track versus same direction rules).

An Example

The functioning of the conflict resolution algorithm can be best illustrated by means of an example.

Three aircraft are following the same track through the airspace at the same flight level (FL390), situated one behind the other at approximately fifteen minute intervals (PAN222, KLM333, and ACA444 in Figure 2). There is also a fourth plane at the next right-way altitude below these three (TWA555 in Figure 2). Soon after the last plane enters the airspace, the leading plane reports a decrease in ground speed of 40 knots. This will cause a loss of longitudinal separation in 12 minutes as the following (second) plane begins to overtake the leading plane.

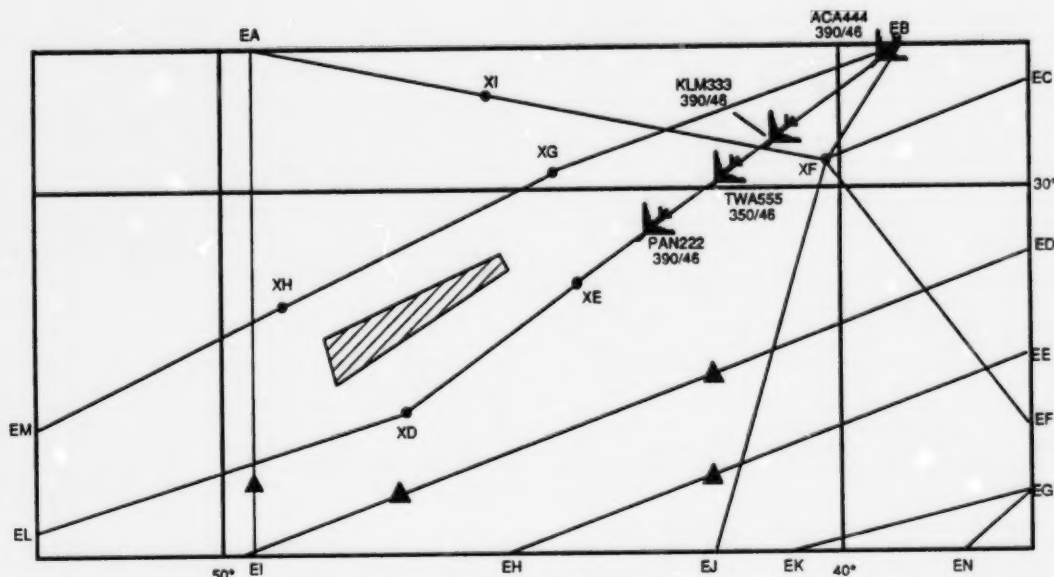


Figure 2: Overtaking Conflict

Conflicts are evaluated based on the two planes immediately involved. For example, although the third plane may eventually overtake the first plane, only the conflict between the first and second planes will be evaluated at this time.

During the initial resolution stage, the conflict is classified as overtaking, with both planes level. This results in the removal of the following strategies from future consideration: all delays (i.e., hold, decrease airspeed, delay arrival) for the leading plane, and vertical detours for either plane.

The remaining strategy set is further reduced by testing plane performance characteristics and airspace criteria. This causes the invalidation of climbing either plane to the next right-way altitude, as the flight level (FL430) is outside of the operating altitude for these aircraft.

Following this, any remaining strategies are computed to determine exact maneuvers, and are tested for any potential conflicts through the airspace. This test will discard solutions which delay the second plane (will induce a conflict with the last plane), and those which descend either plane to the next right-way altitude (will induce a conflict with the plane TWA555 at that altitude [FL350]). Note that, if necessary, the execution of each strategy will be computed to occur at various times.

As no solutions have been found by the end of the initial stage, the algorithm will continue to the relax constraints stage, and attempt more extreme maneuvers of any strategies that have failed due to induced conflicts. This results in the following solutions:

climb or descend either plane to the next wrong-way altitude (FL410 or FL370), and

reroute either plane to XG-XH-EM. (This is a more extreme strategy, as either plane is routed directly from its current position to XG, not following any predefined track.)

Due to a number of solutions found at this stage, the convergence stage is unnecessary. However, consider for a moment that all possible resolution procedures have failed for some reason. In this case, the convergence stage will evaluate all procedures that failed due to induced conflicts. This evaluation will produce only those solutions that resulted in a more desirable airspace. For example, if the original conflict is considered critical then any partial solution which induces a non-critical conflict is acceptable.

The solutions found are ranked and presented in decreasing order to the controller. Out of a maximum of 1, these solutions are ranked fairly low, as they are a result of relaxing constraints. The ranking procedure is based on the following criteria:

- the type of resolution maneuver,
- whether the plane being maneuvered is a pop-up,
- the change in flying time through the airspace,
- the times the plane has been previously moved,
- whether the resolution procedure is a result of the initial stage or the relax constraints stage, and
- the weight category of the plane.

Solutions			Rank
Descend/FL370	PAN222	.54	↑
Climb/FL410	PAN222	.54	↑
Descend/FL370	KLM333	.54	↑
Climb/FL410	KLM333	.54	↑
Reroute	KLM333	.37	↓
Reroute	PAN222	.37	↓

Figure 3: Conflict Solutions

Results

The goals of the project were all achieved. The resolution advice presented by the system seems reasonable under the constraints of the system. The system was successfully tested with a set of scenarios designed to exercise all stages of the algorithm. These scenarios were reviewed by ATC personnel as being suitable tests for this prototype.

When the project to develop CORES began, we completely underestimated the complexity and effort required to develop the conflict prediction mechanism. Although a simpler prediction algorithm would have been sufficient to demonstrate the functionality of the resolution algorithm, it would have detracted from a controller's perception of efficient airspace utilization. Interestingly, the conflict prediction mechanism is a performance bottle-neck. On average when solving a conflict, 75% of the time is spent testing potential resolution procedures, with the remainder selecting and computing the resolution procedures. However, to our knowledge no other automated conflict prediction mechanism intelligently applies as many ATC rules and procedures.

Other performance penalties came from integrating a graphics display system and a simplified simulator with the conflict prediction and resolution systems on the same processor. We found the 386 version of KEE gave particularly poor graphics performance. In addition, the version of Lisp on which 386-KEE is based is not multi-tasking and uses "stop-and-copy" garbage collection. This lack of multi-tasking meant that we were not able to show the three main algorithm stages running at the same time (each stage has to complete before the next can start). The stop-and-copy garbage collection detracted from the overall impact of the system.

To date, scenarios used to force the resolution algorithm to "call the controller for help" have been unrealistic in nature. This confirmed our original thoughts (as reported in [4]) that under the constraints of the current ATC non-radar environment an operational system would rarely, if ever, not be able to produce some form of resolution advice.

Direction for the Future

It has been proposed that the current CORES system be extended for use in general flight planning applications. It could be applied to both the preplanning of flights and to strategic planning for dynamic in-flight adjustments. The advantage of this system is that it could predict controller workloads not only in terms of the number of aircraft per sector but also in terms of the difficulty of possible conflict situations that may arise.

As an extension, the system simulation capabilities could be adapted to evaluate new route structures or changes in procedures in terms of controller workload.

There are difficulties in adapting the CORES system directly to a radar environment. The main reason for these difficulties is the lack of facilities for an automated system to accurately predict aircraft flight paths between approximately 3 and 20 minutes ahead. Currently, for the non-radar environment, CORES predicts flight paths using flight plan information and pilot position reports. In a radar environment, flight paths can be predicted up to about three minutes ahead using only radar returns, and flight plans are only suitable for deriving very long term intent. Neither of these sources are suitable for predicting accurate aircraft trajectories between 3 and 20 minutes ahead. Flight plans in a radar environment are missing critical information. For example, small vectoring adjustments to traffic may be planned by a controller, but are not recorded anywhere for an automated system to access. The introduction of Automatic Dependent Surveillance (ADS) will dramatically change this.

However, before ADS is common place there is a need to investigate automated tactical/emergency conflict resolution advisory for controllers. Unlike the strategic system used in CORES, the overriding constraint of a tactical system is immediate, accurate response. The advice presented must also be in a form which is instantly understandable and useable to a controller. Meeting these requirements will be challenging. Developing a prototype tactical (radar-based) conflict resolution advisory system will be our next major research and development effort.

Summary & Conclusions

The feasibility of an operational conflict resolution advisory system has been demonstrated. Unlike other prototypes which are incomplete or limited to restrictive types of scenarios, CORES provides a vehicle for testing the effectiveness of an automated advisory function under a wide range of circumstances (including traffic patterns which are unrealistic in today's environment). Issues for future study include, how to best integrate this type of system into either current or future operations and how to best present resolution advice in an effective manner.

Acknowledgements

The CORES project was funded by the Transportation Development Centre (TDC) of Transport Canada under the direction of Mr. James Reid, Senior Development Officer. Mr. Pierre Sauve of Air Traffic Rules and Procedures Development, Transport Canada and Mr. Wayne Sherriff of the Research and Experimentation Centre, Transport Canada provided the controller expertise necessary to develop, test, and validate the system.

The direction for extending this work towards both flight planning and tactical radar environment applications was influenced by Mr. Robert Armstrong, Manager of the Research and Experimentation Centre, Mr. Hugh McCallum of the RAMP project office, and Mr. Micha Avni of TDC.

References

- [1] Bowen, B.A., Beaton, J.D., Standen, N.M. "A Conflict Resolution Algorithm for Sparse Air Spaces using a Hybrid Knowledge-Based System", Proceedings of the Canadian Aeronautical and Space Institute, Sixth Canadian Symposium on Navigation, Vancouver, B.C., May 1986.
- [2] Bowen, B.A., Beaton, J.D., Standen, N.M. "Feasibility Demonstrated for Automatic Conflict Resolution", ICAO Bulletin, September 1986, pp. 25-27.
- [3] Bowen, B.A. "An Expert System For Aircraft Conflict Resolution in Dense Airspaces", in Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organization, August 1987, pp. 4-1 to 4-7.
- [4] I.G. Smith, D.G. Bowen, and B.A. Bowen, "CORES: A Knowledge-Based System For Aircraft Conflict Resolution In Non-Radar Airspace," Proceedings of the Canadian Conference On Electrical And Computer Engineering, November 3-4, 1988.

**OTHER PAPERS OF INTEREST/
AUTRE PRÉSENTATIONS D'INTÉRÊT**

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989
SLIDING MODE MOTION CONTROL OF AN AUTONOMOUS VEHICLE

Romano M. DeSantis, Richard Hurteau
Ecole Polytechnique de Montréal
P.O. Box 6079 - Station "A"
Montréal, Canada
H3C 3A7

The modalities and the potential for improvement characterizing the application of recently proposed PI and PID sliding mode controllers to the motion control of electric powered autonomous vehicles are investigated. A specific application to the motion control of a commercial power wheel chair is considered.

1. Introduction

Under nominal and/or stationary operating conditions the usage of linear control strategies in motion controllers for automatically guided vehicles, AGVs, is characterized by a satisfactory dynamic behavior and an adequate steady state disturbance rejection. In the presence of transient operating conditions, model nonlinearities, and large external perturbations (e.g.: Coulomb friction), however, such a usage may also lead to unpredictable, out of specification behavior. While "ad hoc" nonlinear modifications such as the complementing of the linear actions with supplementary signals, the application of optimal and adaptive control [2-4] or the changing of the controller configuration as a function of the state of operation of the vehicle may solve the problem, one is never quite sure about just what kind of modification will do.

The objective of the present paper is to explore the modalities of application and the potential for improvement which might be offered in this respect by the PI and PID sliding mode controllers recently proposed in [5,6]. In view of both their amenability to a simple analytic treatment and their practical importance, our attention will be confined to electric powered AGVs of the type usually found in indoors applications such as automatic office mail distribution, industry asynchronous assembly lines, hospital services, power wheel chairs and similar. The locomotion module adopted in these vehicles typically consists of a cart equipped with two rear co-axial drive wheels and two front pivoting free wheels; each drive wheel is independently powered by a DC servo-motor which is in turn energized by a PWM amplifier; the function of the motion controller is to receive from a sensing module data relative to speed, position and orientation of the vehicle; to provide the input to the PWM amplifiers so as to make these parameters follow a set of pre-assigned values with respect to a given path.

2. Vehicle dynamic model

To compute vehicle speed, position and orientation as a function of the voltages applied to its locomotion module servomotors, it is well to first establish the relation between the angle of rotation of the drive wheels and these voltages. Using the notations of table 1, this relation is described by the state model

$$\dot{x} = Ax + BV + \bar{B} \text{PERT}$$

where

$$x' \triangleq [\dot{\theta}_1 \dot{\theta}_2 x_c y_c \phi]$$

$$v' \triangleq [V_1 V_2], \quad \text{PERT}' \triangleq [\text{PERT}_1 \text{PERT}_2]$$

$$A \triangleq \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \end{bmatrix} \quad B \triangleq \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$\bar{B} \triangleq \begin{bmatrix} b'_{11} & b'_{12} \\ b'_{21} & b'_{22} \end{bmatrix}$$

and

$$a_{11} \triangleq - \left[\left(\frac{L^2}{4J} + \frac{1}{M} \right) (f_1 + \frac{K_{b1} K_{\phi 1}}{R_{a1}} \frac{1}{R_{a1}}) \right]$$

$$a_{12} \triangleq \left[\left(\frac{L^2}{4J} - \frac{1}{M} \right) (f_2 + \frac{K_{b2} K_{\phi 2}}{R_{a2}} \frac{1}{R_{a1} R_{a2}}) \right]$$

$$a_{21} \triangleq \left[\left(\frac{L^2}{4J} - \frac{1}{M} \right) (f_1 + \frac{K_{b1} K_{\phi 1}}{R_{a1}} \frac{1}{R_{a1} R_{a2}}) \right]$$

$$a_{22} \triangleq - \left[\left(\frac{L^2}{4J} + \frac{1}{M} \right) (f_2 + \frac{K_{b2} K_{\phi 2}}{R_{a2}} \frac{1}{R_{a2}}) \right]$$

$$a_{31} \triangleq \frac{R_{a1}}{2} \sin \phi \quad a_{32} \triangleq \frac{R_{a2}}{2} \sin \phi$$

$$a_{41} \triangleq \frac{R_{a2}}{2} \cos \phi \quad a_{42} \triangleq \frac{R_{a2}}{2} \cos \phi$$

$$a_{51} \triangleq - \frac{R_{a1}}{L} \quad a_{52} \triangleq \frac{R_{a2}}{L}$$

$$b_{11} \triangleq \left(\frac{L^2}{4J} + \frac{1}{M} \right) \frac{K_{\phi 1}}{R_{a1} R_{a1}} \quad b_{12} \triangleq \left(\frac{1}{M} - \frac{L^2}{4J} \right) \frac{K_{\phi 2}}{R_{a2} R_{a1} R_{a2}}$$

$$b_{21} \triangleq \left(\frac{1}{M} - \frac{L^2}{4J} \right) \frac{K_{\phi 1}}{R_{a1} R_{a1} R_{a2}} \quad b_{22} \triangleq \left(\frac{L^2}{4J} + \frac{1}{M} \right) \frac{K_{\phi 2}}{R_{a2} R_{a2}}$$

$$b'_{11} \triangleq - \left(\frac{L^2}{4J} + \frac{1}{M} \right) \frac{1}{R_{a1}} \quad b'_{12} \triangleq - \left(\frac{1}{M} - \frac{L^2}{4J} \right) \frac{1}{R_{a1}}$$

$$b'_{21} \triangleq - \left(\frac{1}{M} - \frac{L^2}{4J} \right) \frac{1}{R_{a2}} \quad b'_{22} \triangleq - \left(\frac{1}{M} + \frac{L^2}{4J} \right) \frac{1}{R_{a2}}$$

From these equations one obtains the following dynamic model

$$\text{SPEED} = \frac{R_{a2} \dot{\theta}_2 + R_{a1} \dot{\theta}_1}{2}$$

$$\text{SPEED} = - \frac{\text{SPEED}}{\tau_{\text{SPEED}}} + \frac{K_{\text{SPEED}}}{\tau_{\text{SPEED}}} (u_0 + \xi_1)$$

$$\dot{\text{ORIENT}} = - \frac{\text{ORIENT}}{\text{FORIENT}} + \frac{K_{\text{ORIENT}}}{\text{FORIENT}} (\Delta u + \xi_2)$$

$$\dot{x}_c = \text{SPEED} \sin(\text{ORIENT}) \quad \dot{y}_c = \text{SPEED} \cos(\text{ORIENT})$$

$$\text{with: } u_H \triangleq \frac{V_1 + V_2}{2} \quad \Delta u \triangleq \frac{V_2 - V_1}{2}$$

and:

$$r_{\text{SPEED}} \triangleq \left[\frac{2}{MR_0^2} \left(1 + \frac{K_b K_\phi}{R_0} \right) \right]^{-1}$$

$$K_{\text{SPEED}} \triangleq \frac{R_0}{K_b \left(1 + \frac{R_0 R_1^2}{K_b K_\phi} \right)}$$

$$\text{FORIENT} \triangleq \left[\frac{2L^2}{4J R_0^2} \left(1 + \frac{K_b K_\phi}{R_0} \right) \right]^{-1}$$

$$K_{\text{ORIENT}} \triangleq \frac{2R_0/L}{K_b \left(1 + \frac{R_0 R_1^2}{K_b K_\phi} \right)}$$

3. Motion Controller Design

The structure of the above model suggests that the motion controller be viewed as generating control actions u_H and Δu rather than V_1 and V_2 . It also suggests that the controller be given by the ensemble of two decentralized components: - a translational speed controller to generate the control action u_H required to reduce to zero the difference between actual and desired translational speed of the vehicle; - a position/orientation controller to generate the control action Δu required to reduce to zero the difference between actual and desired position and orientation of the vehicle.

The design of these two components will be considered in the special case where the path is a straight line, and the required lateral and orientation offsets are null.

3.1 Translational Speed Controller.

Following the PI/sliding mode development in [5], the control action u_H generated by this controller is viewed as given by the contribution of three components

$$u_H = u_{H1} + u_{H2} + u_{H3}$$

where: u_{H1} is the feedforward component required to obtain the desired behavior specifications under nominal conditions and in the absence of perturbations; u_{H2} is the linear feedback component provided by the classical PI controller; u_{H3} is an on/off switching feedback component which reinforces the intended effect of u_{H1} and u_{H2} by neutralizing the influence of parameter variations, perturbations and model uncertainties.

By specifying that the vehicle follow the desired translational speed trajectory according to an error dynamics described by:

$$\text{SPEED\_OFFSET} + a_{11} \cdot \text{SPEED\_OFFSET} = 0$$

with a_{11} a given time constant, the values of u_{H1} , u_{H2} and u_{H3} are computed via the formulas

$$u_{H1} = \frac{r_{\text{SPEED}}}{K_{\text{SPEED}}} \text{DES\_SPEED} + \frac{1}{K_{\text{SPEED}}} \text{DES\_SPEED}$$

$$u_{H2} = K_{11} \cdot \text{SPEED\_OFFSET} + K_{12} \int \text{SPEED\_OFFSET} dt$$

$$u_{H3} = M \cdot \text{SIGN}[\text{slm}_1] \text{ if } |\text{slm}_1| > \epsilon_1, = 0 \text{ otherwise}$$

where

$$\text{SPEED\_OFFSET} \triangleq \text{DES\_SPEED} - \text{SPEED}$$

$$\text{slm}_1 \triangleq a_{11} \text{SPEED\_OFFSET} + \int \text{SPEED\_OFFSET} dt$$

and DES_SPEED: represents the desired translational speed; K_{11} , K_{12} , a_{11} , M and ϵ_1 are the controllers parameters.

The values of the controller parameters are chosen by following the procedure exposed in [5]. Indicating with p_{11} and p_{12} pre-assigned poles characterizing the desired dynamics of the speed offset, one has

$$K_{11} = - [(p_{11} + p_{12}) r_{\text{SPEED}} + 1] / K_{\text{SPEED}}$$

$$K_{12} = p_{11} p_{12} r_{\text{SPEED}} / K_{\text{SPEED}}$$

$$a_{11} = -1/p_{11}$$

Positive M_1 is chosen to be as large as possible up to the maximum expected value of the input equivalent perturbation ξ_1 . Positive ϵ_1 is chosen so that under "small" speed offsets the controller feedback action coincides with that which would be provided by a classical PI; under larger speed offsets, the more energetic action of the switching component enters into play.

3.2. Position/Orientation Controller

The lateral offset of the vehicle is given by the y coordinate of the origin of the vehicle frame, IRv, with respect to the path dependent frame, IRs. This implies

$$\text{LAT\_OFFSET} = n_1 x_c + n_2 y_c$$

where $\Omega = (n_1, n_2)$ is the unit vector perpendicular to the path. The orientation offset is simply given by

$$\text{ORIENT\_OFFSET} = \text{DES\_ORIENT} - \text{ORIENT}$$

where DES_ORIENT: represents the desired orientation.

By specifying that the vehicle follow the assigned path with an error dynamics described by:

$$\text{LAT\_OFFSET} + a_{21} \cdot \text{ORIENT\_OFFSET} + a_{22} \cdot \text{ORIENT\_OFFSET} = 0$$

with a_{21} and a_{22} given parameter values, the design of the position/orientation controller may be implemented by adopting the PID/sliding mode structure of the type considered in [8]. In analogy with the translational speed controller case, the control action Δu may once again be viewed as given by the contribution of three components: Δu_1 is the

feed-forward component required to obtain the desired performance under nominal conditions and in the absence of perturbations; Δu_2 is the linear feedback component provided by the classical PID controller; Δu_3 is an on/off switching feed back component with the objective of reinforcing the intended effect of Δu_1 and Δu_2 by neutralizing the influence of external perturbations, parameter variations and model uncertainties.

Δu_1 , Δu_2 and Δu_3 are computed by

$$\Delta u_1 = \frac{\text{DES\_ORIENT}}{K_{\text{ORIENT}}} + \frac{\text{DES\_ORIENT}}{K_{\text{ORIENT}}}$$

$$\Delta u_2 = K_{21} \text{ORIENT\_OFFSET} + K_{22} \text{ORIENT\_OFFSET} + \frac{K_{23}}{\text{DES\_SPEED}} \text{LAT\_OFFSET}$$

$$\Delta u_3 = M_2 \text{SIGN}[\text{sim}_2] \text{ if } |\text{sim}_2| > \epsilon_2, = 0 \text{ otherwise}$$

where $\text{sim}_2 = \frac{\Delta}{\epsilon}$

$$\frac{\text{LAT\_OFFSET}}{\text{DES\_SPEED}} + a_{21} \text{ORIENT\_OFFSET} + a_{22} \text{ORIENT\_OFFSET}$$

and K_{21} , K_{22} , K_{23} , a_{21} , a_{22} , M_2 and ϵ_2 are the controller's parameters.

The values of these parameters are once again chosen by following the procedure exposed in [6], indicating with p_{21} , p_{22} and p_{23} pre-assigned poles characterizing the desired dynamics of position and orientation offsets, one has

$$K_{21} = (p_{21} p_{22} + p_{21} p_{23} + p_{22} p_{23}) \frac{\text{ORIENT}}{K_{\text{ORIENT}}}$$

$$K_{22} = -((p_{21} + p_{22} + p_{23}) \text{ORIENT} + 1) / K_{\text{ORIENT}}$$

$$K_{23} = -p_{21} p_{22} p_{23} \text{ORIENT} / K_{\text{ORIENT}}$$

$$a_{22} = \frac{1}{p_{21} p_{22}} \quad a_{21} = -a_{22}(p_{21} + p_{22})$$

In analogy with the translational speed controller case, the value of positive M_2 is chosen so as to be as large as possible up to the maximum expected value of the input equivalent perturbation ϵ_2 . The value of positive ϵ_2 is chosen so that under "small" orientation and lateral offsets the controller feedback action coincides with that provided by a classical PID; under larger speed offsets, the more energetic action of the switching component enters into play.

4. An application Example

Consider the design of a motion controller for the power wheel chair model 8750 by Fortress Engineering Co. (a Montreal based company). The commercial version of the locomotion module of this chair is characterized precisely by the geometry, the components, and the dynamic model discussed in the previous sections. The principle of operation of its controller is quite primitive: to increase the translational speed, the operator increases V_1 and V_2 by an equal amount by pushing forward a joystick; to change orientation, say in the CCW direction, V_2 is made bigger than V_1 by pushing the joystick to the left side; to decrease translational speed or change orientation in the CW direction the opposite is done. When the joystick is in its rest position, then the drive wheels are blocked by the action of a powerful pair of electro-magnetic brakes.

A large variety of operating modes may be envisioned where the performance of such a controller might be substantially improved. In the case where the wheel chair has to follow a desired path with a pre-assigned speed and lateral offset, (cruising mode), such an improvement may be obtained by applying "verbatim" the development of the previous sections. From a hardware and instrumentation point of view, this can be done by adopting a number of practical approaches. For the sake of concreteness, we consider the specific solution adopted in our laboratory: the controller is implemented by using an IBM PC micro-computer equipped with a LabMaster input/output interface; odometric measurements are obtained by installing optical encoders on each drive wheel axis; inertial measurements are obtained by processing the image of a line painted on the ceiling as captured by a 2-D camera mounted on the vehicle; the vehicle speed is directly computed from the odometric measurements; the lateral and orientation offsets are computed via a hybrid scheme combining inertial and odometric measurements; the operator interface is provided by the PC keyboard.

By taking into account the Fortress locomotion module parameter values reported in table 2, the motion controller design unfolds as follows:

Speed controller: By requiring $p_{11} = p_{12} = -1$, one obtains

$$K_{11} = -((p_{11} + p_{12}) \text{rspeed} + 1) / K_{\text{SPEED}} = 10$$

$$K_{12} = p_{11} p_{12} \text{rspeed} / K_{\text{SPEED}} = 20$$

$$a_{11} = -1/p_{11} = 0.5 \text{ sec}$$

Since a strong Coulomb friction on the forward motion is hypothesized, the values of M_1 and ϵ_1 are selected as equal to $M_1 = 100 \text{ volt}$ $\epsilon_1 = 0.01 \text{ rad}$.

Position/orientation controller: By requiring the poles $p_{21} = p_{22} = p_{23} = -2$ one obtains

$$K_{21} = (p_{21} p_{22} + p_{21} p_{23} + p_{22} p_{23}) \frac{\text{ORIENT}}{K_{\text{ORIENT}}} = 15$$

$$K_{22} = -((p_{21} + p_{22} + p_{23}) \text{ORIENT} + 1) / K_{\text{ORIENT}} = 5$$

$$K_{23} = -p_{21} p_{22} p_{23} \text{ORIENT} / K_{\text{ORIENT}} = 0$$

$$a_{22} = \frac{1}{p_{21} p_{22}} = 0.25 \text{ sec}^2 \quad a_{21} = -a_{22}(p_{21} + p_{22}) = 1 \text{ sec}$$

Hypothesizing a moderate perturbation uncertainty in regard to steering, the values of M_2 and ϵ_2 are selected as equal to zero.

The behavior of the motion controller is illustrated by considering a simulation experiment whereby, in addition to the above values, the following elements are taken into account: - the influence of a strong forward motion Coulomb friction (with a value equivalent to 100 Volts); - a desired translational speed equal to 5 m/s; - initial offsets: lateral = -0.5 m, orientation = -15 degrees, speed = 5 m/s; - sampling period : 0.005 sec.

Figure 1 illustrates the influence of the Coulomb friction on the dynamic behavior of the vehicle in the case where the classical PI, ($M_1 = 0$), is used in the speed controller. At the beginning of the experiment the Coulomb friction has a stronger value than the proportional control torque applied to the wheels: the vehicle remains then stuck until the integral control action becomes sufficiently strong to overcome the friction. Figure 2 is to the effect

that the same Coulomb friction has almost no influence on the vehicle dynamics when the classical PI is modified into a PI/sim : the response of the system with or without the presence of friction is essentially the same and corresponds to that obtained in the case of the classical PI under no friction. One notices, on the other hand, that the on-off action of the controller nonlinear switching component has now induced the presence of high frequency oscillations in the vehicle dynamics. In actual experimental tests these oscillations should, however, be easily damped by the low pass nature of the locomotion module.

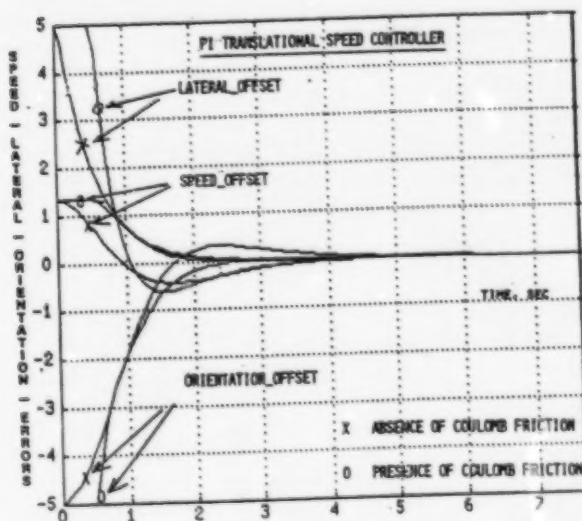


Figure 1: Effect of Coulomb friction in the classical PI speed controller

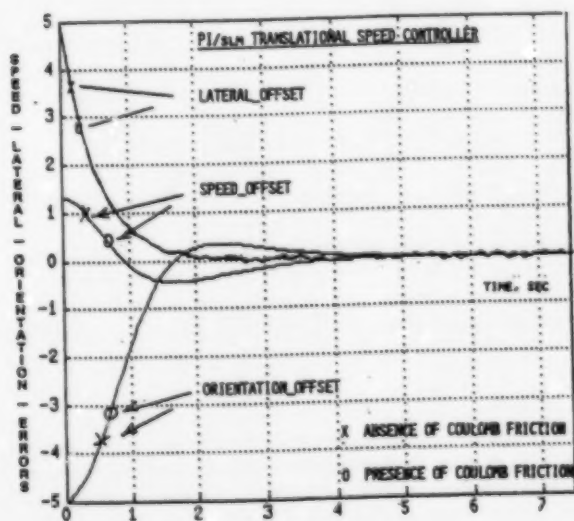


Figure 2: Effect of Coulomb friction in the PI/Sim speed controller

Table 1: Physical interpretation of the symbols

Distance between the drive wheels L
 Drive wheel radius R_{a1}, R_{a2}
 Drive wheel viscous damping coefficient f_1, f_2
 Voltage applied to the servo-motors V_1, V_2
 Torque applied by the servo-motor on the drive wheel axle Γ_1, Γ_2
 Propulsion force applied to the drive wheel F_1, F_2
 Perturbation applied to the drive wheel P_{ERT1}, P_{ERT2}
 Servomotor internal resistance R_{s1}, R_{s2}
 Servomotor Voltage constant K_{b1}, K_{b2}
 Servomotor Torque constant $K_{\phi1}, K_{\phi2}$
 Mass of the Vehicle M
 Moment of Inertia of the Vehicle w/r to its z axis J
 Angular Position of the drive wheel θ_1, θ_2
 Vehicle frame origin coordinates w/r to the inertial system (x_e, y_e)
 Vehicle translation speed, $SPEED$
 Vehicle orientation ϕ , $ORIENT$
 Matrices of the (nonlinear) state model $A, B, \bar{B}$
 Vehicle lateral offset LAT_OFFSET
 Vehicle orientation offset $ORIENT\_OFFSET$
 Speed controller action $u_M = (V_1 + V_2)/2$
 Position/orientation controller action $\Delta u = (V_2 - V_1)/2$
 Vehicle speed gain and time constant K_{SPEED}, τ_{SPEED}
 Vehicle orientation gain and time constant $K_{ORIENT}, \tau_{ORIENT}$
 Input equivalent perturbation ξ_1, ξ_2
 Motion Controller Linear Component Gains K_{ij}
 Motion Controller Switching Component Parameters a_{ij}, M_i, e_i, ξ_i

Table 2: Fortress Locomotion Module (typical) Parameter Values

Distance between the drive wheels $L = 0.5$ m
 Drive wheel radius $R_{a1} = R_{a2} = 0.12$ m
 Mass of the Vehicle $M = 200$ Kg
 Moment of Inertia of the Vehicle w/r to its yaw axis $J = 12.5$ Kg $\cdot$ m ^{2}
 Vehicle translational speed gain $K_{SPEED} = .1m/volt\cdot sec$
 Vehicle translational time constant $\tau_{SPEED} = .5$ sec
 Vehicle orientation gain $K_{ORIENT} = .4$ rad/volt $\cdot$ sec
 Vehicle time constant $\tau_{ORIENT} = .5$ sec

References

- [1] Anonymous, *DC Motors Speed Controls Servo Systems*, Engineering Handbook by Electro-Craft Corp., Pergamon Press, 1977.
- [2] Ellis, J.R., *Vehicle Dynamics*, London: London, Business Books, 1969.
- [3] Nisouger, R.L., Wormley, D.N., Dynamic Performance of Automated Guideway Transit Vehicles with Dual-Axis Steering, *IEEE Trans. on Vehicular Technology*, vol. VT-28, no. 1, February 1979.
- [4] Fenton, R.E., Selim, I., On the Optimal Design of an Automotive Lateral Controller, *IEEE Trans. on Vehicular Technology*, vol. 37, no. 2, May 1988.
- [5] De Santis, R.M., On PI and PID/Sliding Mode Controllers, *Advances in Communication and Control*, Springer Verlag, December 1989.
- [6] De Santis, R.M., An Adaptive PI/Sliding Mode Controller for a Speed Drive, *ASME Journal of Dynamic Systems, Measurement and Control*, December 1989.
- [7] Rabemanantsoa, M., De Santis, R.M., Système d'Autopilotage d'un Robot Mobile, RTXX, Ecole Polytechnique de Montréal (to appear).
- [8] Vignola, P.-L., Hurteau, R., Identification des paramètres du modèle d'un robot mobile, RTXX, Ecole Polytechnique de Montréal, (to appear).

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

Coherent Detection of Interleaved Trellis Encoded CPFSK on Shadowed Mobile Satellite Channels

Ron W. Kerr and P.J. McLane

Department of Electrical Engineering
Queen's University at Kingston

ABSTRACT—To date, modems for use on fading land Mobile Satellite (MSAT) channel, which employ coherent detection and interleaving, have used forms of *Phase Shift Keying* (PSK) for the modulation scheme. The objective of this paper is to present a receiver that utilizes Trellis Coded *Continuous Phase Frequency Shift Keying* (CPFSK) with coherent detection and convolutional interleaving of the coded symbols. A new decoding algorithm that copes with the interleaving is presented and uses both the modulation trellis and the code trellis. One earlier consideration considered Trellis Coded CPFSK with differential detection where decoding is based solely on the code trellis. The performance results are obtained by Monte-Carlo simulation and are partially verified through analysis. The fading model is Ricean, but the Line-of-Sight (LOS) component is subjected to a log-normal transformation that represents attenuations due to foliage, that will be referred to as shadowing. The main application for the receiver is low-speed-data transmission which requires a low bit error rate (BER) and has no delay constraints as in digital speech transmission.¹

I. INTRODUCTION

The Mobile Satellite (MSAT) Communication channel is both power and band-limited, thus the need arises for a power and bandwidth efficient coding/modulation scheme. Trellis coded modulation (TCM) [1-3] combined with Continuous Phase Modulations was chosen for this investigation because of CPM's low spectral occupancy and constant envelope and TCM's ability to improve the system performance without increasing power on bandwidth requirements.

Convolutional interleaving of the code symbols was used, as in [5] to combat the effects of fading. The problem arises when interleaving a CPM signal since the phase information, which is usually utilized in the decoding, is destroyed by the de-interleaving process. In this paper a coherent detection scheme for CPFSK with interleaving for use for low-speed mobile satellite data service is presented which utilizes the modulator and coder trellis in the decoding.

A differential detection scheme was presented in [6] which used block interleaving of the code symbols. This system was intended for digital speech service at 4800 bps.

II. SYSTEM DESCRIPTION

The block diagram for the model for the digital communication system considered in this paper is shown in Figure 1. Continuous Phase Frequency Shift Keying (CPFSK) is used for the modulation and is a subset of CPM [13]. The signal for CPFSK can be written

as:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos \left(\omega_c t + \frac{\alpha_i \pi h (t - iT)}{T} + \phi_i \right), \quad (1)$$

$$iT \leq t \leq (i+1)T$$

where

ϕ_i is the phase at the start of the symbol interval,

ω_c is the radian carrier frequency,

h is the modulation index,

α_i is the symbol sent in the i^{th} interval,

T is the symbol interval time,

and

E is the symbol energy.

For the system in our study, we set $h = \frac{1}{4}$. The uncoded modulation is binary CPFSK where $\alpha_i = \pm 1$ are the modulation symbols. In the coded case depicted in Figure 1, a rate = $\frac{1}{2}$ convolutional encoder is followed by a mapping rule, and thus there are four modulation symbols, namely $\alpha_i = \{\pm 1, \pm 3\}$: that is the modulation is 4-level CPFSK. The encoders used were (13, 4) for the 8 state code and (23, 14) for the 16 state code, both used the natural mapping rule: (13,4) octal represents top location.

The signal space method was used for our simulations; an explanation can be found in [8]. With this method, a basis of orthonormal functions found from the modulation signals, the channel signals and AWGN can then be expressed as vectors on this basis. The received signal during the i^{th} baud interval in a fading environment with ideal coherent detection using this vector notation is given by

$$\underline{r}_i = \rho_i \underline{g}(\phi_i, \alpha_i) + \underline{n}_i, \quad (2)$$

where ρ_i is the (normalized) random fading amplitude of the signal envelope in the i^{th} interval. In this notation, $\underline{n}_i$ is a vector of gaussian variables with zero mean and variance of $\frac{N_0}{2}$, $\underline{g}(\phi_i, \alpha_i)$ is the signal sent in the i^{th} interval, ϕ_i was the initial phase state of the modulator and α_i was the symbol sent. This simulation method and fading model were used for work done in [10] on Multi-h modulations, and in [14] for work on CPFSK.

III. FADING PROCESS

The generation of the Ricean fading process with line-of sight shadowing is shown in Figure 2. The Gaussian Random Variable (GRV) was filtered using a third order Butterworth Filter, which gave the best match to experimental results in [4]. This method was used in [9] where PSK modulation was considered and in [10] where Multi-h modulation was used.

¹ This research was supported by the Canadian Department of Communications, the Natural Sciences and Engineering Research Council of Canada and the Telecommunications Research Institute of Ontario (TRIO)

The fading amplitude is represented by the magnitude of the complex random variable for each symbol interval. Under the assumption of perfect coherence, only amplitude fades affect the system. The fade is assumed to be constant over the entire symbol interval. This follows earlier work in [9].

3.1 Fading Model Parameters

The parameters for the fading model are from [4] and correspond to three different degrees of shadowing of the LOS component, namely light, average and heavy. Light and average shadowing were the cases considered in this paper. The model was developed for the Canadian MSAT system where fading is more severe than in the United States, since the angle of elevation to the geosynchronous satellite is lower in Canada.

There are three parameters required to define the type of shadowing. The parameters that are given in Table 1 are based on experimental measurements and correspond to a normalized signal envelope of unity. In the table, u_0 and $\sqrt{d_0}$ are the mean and the standard deviation, respectively, of u where the log-normal transformation of the LOS component is e^u .

Separate third-order Butterworth filters provide the time variation for the attenuated LOS component and the scatter component. The bandwidth of these filters are taken to be the same. Although the fading process is time varying, it can be considered flat fading relative to the modulation process. The fading bandwidth, normalized to the symbol rate, $1/T$, is $B_d T$ where B_d is the Doppler shift of the carrier due to the vehicle motion.

IV. DECODING ALGORITHM

It was shown in [7,15] that observing over more than one interval for CPFSK could improve the error performance of the system. The algorithm presented here uses the modulation trellis, as shown in Figure 3, to observe the CPFSK signals over two intervals via signal processing. More than two intervals are possible, but only two were considered in this investigation. The information from the two symbol intervals was used in the calculations of the symbol metrics, instead of making a decision on the bit sent as was done in [7,15]. This method allows soft-decision Viterbi decoding after the deinterleaver and takes advantage of the continuous phase and half-reachable property of the CPFSK modulation.

The algorithm uses the fact that the initial phase and symbol sent in the i^{th} interval defines the initial phase for the $i+1^{st}$ interval. For example, for 4 Level CPFSK the relationship between the signal in the i^{th} interval and the $i+1^{st}$ interval's initial phase is

$$\phi_{i+1} = (\phi_i + \alpha_i) \text{ Mod } 8. \quad (3)$$

Therefore, a contribution of the symbol metrics in the i^{th} interval can be made by finding the minimum metric from the set of expected signals with an initial phase ϕ_{i+1} where ϕ_{i+1} is defined in Equation 3.

The calculation over two intervals increases the decoding delay by one, but realizes again of about 2 dB over calculating the metrics on a symbol-by-symbol basis from our simulations.

The formula for calculating the metrics is as follows

$$\text{Metric}(\alpha_i) = \min_{\phi_i} \left[\|\mathbf{r}_i - \mathbf{s}(\phi_i, \alpha_i)\|^2 + \|\mathbf{r}_{i+1} - \mathbf{s}(\phi_{i+1}, \alpha_{i+1})\|^2 \right]$$

$$\text{where } \phi_{i+1} = (\phi_i + \alpha_i) \text{ Mod } 8. \quad (4)$$

Degree of Shadowing			
	Light	Average	Heavy
σ^2	0.158	0.126	0.0631
u_0	0.115	-0.115	-3.91
$\sqrt{d_0}$	0.115	0.161	0.806

Table 1 Parameters for various degrees of shadowing

Basically, the relationship in (3) assigns the minimum Euclidean distance between the received signals and the set of all possible transmitted sequences over two intervals, where α_i is the first symbol to the metric for α_i .

A reduction in computation was possible when implementing the above equation by recognizing that the second term will be a constant for a given state. Thus the minimum metric from the second state can be added to the first term in (4) so (4) can be replaced with

$$\text{Metric}(\alpha_i) = \min_{\phi_i} \left[\|\mathbf{r}_i - \mathbf{s}(\phi_i, \alpha_i)\|^2 + \min_{\alpha_{i+1}} \|\mathbf{r}_{i+1} - \mathbf{s}(\phi_{i+1}, \alpha_{i+1})\|^2 \right] \quad (5)$$

$$\text{where } \phi_{i+1} = (\phi_i + \alpha_i) \text{ Mod } 8$$

$$\alpha_{i+1} = \{\pm 1, \pm 3\}$$

V. INTERLEAVING/DEINTERLEAVING

As was mentioned earlier, interleaving is a method which transforms a burst error channel into a channel with independent errors by re-ordering the symbol sequence. A convolutional interleaver was placed between the trellis encoder and the CPFSK modulator, this is similar to the design in [6], but in [6] a block interleaver was used. Since the interleaver scrambles the symbols, the modulator receives a symbol stream of uncorrelated symbols and effectively the resulting modulation across the channel is uncoded 4-level CPFSK. At the receiver, the deinterleaver deinterleaves the metrics calculated for each symbol by the metric computer. This structure allows soft-decision decoding using a Viterbi decoder after the deinterleaver. The deinterleaver was implemented in the simulations with each delay register, as shown in Figure 4, capable of storing a four element array that contained the symbol metrics.

VI. ESTIMATION OF SYSTEM PERFORMANCE

6.1 No interleaving

If the fading is slow enough to be considered constant over the number of intervals equal to the minimum error event, then for coherent detection with a Gaussian metric, the average bit error probability is asymptotically estimated by

$$P_b \approx C_1 \exp \left[\frac{-\rho^2 d_{free}^2 \frac{E_b}{N_0}}{2} \right] \quad (6)$$

where C_1 is a constant dependent on the distance structure of the code and

$$d_{free}^2 = \min_{\mathbf{x}, \hat{\mathbf{x}} \in C} \sum_{n \in \eta} |x_n - \hat{x}_n|^2$$

The overbar denotes the averaging of the term over the Ricean probability density. The Light Shadowed channel as shown in [4,5] can be approximated by the K=5 dB Ricean channel. The Ricean probability density is

$$P(\rho) = \begin{cases} 2\rho(1+K) \exp[-K - \rho^2(1+K)] I_0(2\rho\sqrt{K(1+K)}) & \rho \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

where $I_0(x)$ is the modified Bessel function. Doing the averaging results in

$$P_b \approx C_1 \frac{1+K}{1+K+C_2} \exp\left[\frac{-K C_2}{1+K+C_2}\right] \quad (8)$$

where $C_2 = d_{free}^2 \frac{E_s}{4N_0}$.

This analysis can be found in [11]. The plot for the non-interleaved system is shown in Figure 5, and shows that the simulations agree with the theory.

6.2 Ideal Interleaving/De-interleaving

Assuming that the interleaving/de-interleaving of the coded symbols results in a memory-less channel, the fading is independent between adjacent symbol intervals in the Viterbi decoder. This analysis is also presented in [11]. From [11] the pairwise error probability with CSI and a Gaussian metric over the Ricean channel is given by:

$$P(x \rightarrow \hat{x}) \geq \prod_{n \in \eta} \frac{1+K}{1+K+\beta_n \gamma} \exp\left\{\frac{-K \beta_n \gamma}{1+K+\beta_n \gamma}\right\} \quad (9)$$

where $\beta_n = |x_n - \hat{x}_n|^2$ and $\gamma = \frac{E_s}{4N_0}$.

The approximation for the system's average bit error performance at very high SNR is given by:

$$P_b \approx C_1 \left[\frac{(1+K) \exp(-K)}{E_s/N_0} \right]^L \quad (10)$$

where $C_1 = C \prod_{n \in \eta} \beta_n^{4L}$

where L is the length of the shortest error event and is defined by the number of non-zero pairwise symbol distances in the error path and C_1 is a constant dependent on the distance structure of the code.

The important point, as was stated in [11], is that P_b is inversely proportional to $(E_s/N_0)^L$ when the symbols along the error path are independently faded. When no interleaving is employed P_b is inverse linear to E_s/N_0 since the length of the shortest error event is one. This results in (9) was used to estimate the behavior of the system although for the results presented, no CSI was used.

Equations (9) and (10) were used to estimate the performance of the (23,14) code on a fast fading channel with ideal interleaving. The estimate and the simulation results of the code on a fast fading channel with a large interleaving degree to approximate ideal interleaving are presented in Figure 6. As we stated earlier the Light shadowed channel can be approximated by a K=5dB Ricean channel [4,5]. As can be seen the estimate is close to the simulated performance on this channel.

6.3 Computer Simulations

These results were obtained from Monte-Carlo simulations using the fading model described above. The fading bandwidth, which is normalized to the symbol rate, models the effect of the vehicle motion. For example, the L-band frequencies with a 1200 sym/sec rate, the fading bandwidth of 0.01, 0.05, and 0.1 represent speeds of about 10, 50 and 100 Km/hr respectively.

The method of determining validity of the simulation points was to run a simulation with a given number of trials and then run a second one with a greatly increased number of trials; if the two results were 10 percent of each other, the point was accepted. If the point failed, then the number of trials was increased again until the results were within 10 percent.

The system described in this paper was simulated with convolutional encoders of 4,8, and 16 states on various fading channels. A comparison of these systems on a slow fading light shadowed channel are plotted against a non-interleaved code that has 16 states in the receiver trellis in Figure 7. This Figure shows that gain due to the interleaving can be achieved even with a 4 state code. Figure 7 also shows that coding gain can be realized in a slow fading environment. The degree of interleaving was set to 18 to ensure that the fading along the error path was independent.

The non-interleaved system is (1,2) Natural Mapped h=1/4 CPFSK and was used for comparison on the various channels since its receiver trellis has 16 states; the same number as the interleaved system with a 16 state encoder. This comparison was chosen since the receiver trellis for both non-interleaved and interleaved systems have the same amount of memory. The 8 state interleaved system is also shown in the following Figures to compare its performance against the 16 state code.

In figures 8-10, the results for the light shadowed channel for the 8 and 16 state codes are presented with various interleaving degrees and normalized bandwidths with the non-interleaved system for comparison.

In Figure 8, shows a comparison between the 8 and 16 state codes. The 8 state code outperformed the 16-state when the degree of interleaving was reduced to 10. This may be a result of the fades not being independent along the error path for the 16-state code. The 16-state code has long error events, since there may exist symbols with a zero pair-wise distance to the symbol in the correct path. Therefore, in such a long error event, a small interleaving degree may introduce correlated fades.

The interleaving degree was changed for the 8 and 16-state codes in the presence of fast fading (BW=.05 or Vehicle speed at 50 Km/hr) in Figure 9. From this Figure it can be shown that a smaller interleaving degree (hence, a smaller decoding delay) can still achieve gains over the non-interleaved system and the two codes perform similarly.

Figure 10 shows the performance of the systems on the light shadowed channel with BW=0.1. The results for this case are similar to those in the previous two Figures; interleaving gains are achievable (although smaller for this case)

Similar results are seen for the average shadowed channel case. Results for this channel for fading bandwidths of BW=0.01 and BW=0.05 are shown in Figures 11-12, respectively.

Results for the 16-state code on the K=10dB Ricean Channel for the same fading bandwidths as the average shadowed case are shown in Figures 13-14. These results show that the gains on this channel are low as compared to the non-interleaved case,

thus interleaved CPFSK on the K=10dB Ricean channel is not as beneficial as on the Canadian MSAT channel.

The results are summarized in Tables 2-4. The fade margins are calculated relative to MSK on the AWGN channel (i.e. $Q(\sqrt{2E_b/N_0})$). The gains are calculated by taking the difference between the SNR in dB of the non-interleaved and interleaved systems at the intercept of $P_b = 10^{-3}$ and $P_b = 10^{-4}$. If the data did not cross these lines the data was extrapolated linearly to the line and this intercept was used in the estimation of the gain.

VII. DISCUSSION AND CONCLUSIONS

Analyzing the trellis representations of the 8 state code and the 16 code used, it was found that the shortest error event for both codes was four symbols long. It is possible for symbols with zero pair-wise distances from the symbols in the correct path, to occur in the error path for the 16 state code. Since both the codes' shortest error event has four non-zero pairwise symbol distances, both should vary inversely with respect to $(\bar{E}_s/N_0)^4$ with ideal interleaving. In the simulations on the fast fading light shadowed channel the two codes performed similarly with high interleaving degree and as shown in Figure 9 and compares well with theory. For the 16 state code, it is possible that long (baud interval length) error events have a smaller d_{free} than shorter error events because of the symbols that make no contribution to the cumulative distance of the error event. This property may cause error bursts even when the fading is independent. This is currently being examined.

The spectral occupancy of CPFSK was one of the major factors in the investigation. The effective modulation across the channel is uncoded 4 level CPFSK, since the code symbols are interleaved before the modulator, thus the symbols are uncorrelated at the modulator. Thus, the system has the same spectral characteristics as uncoded 4 level CPFSK with $h=1/4$.

The results show that interleaving CPFSK and decoding as described above, gives a significant error performance improvement over the 'non-interleaved' CPFSK system, with the same number of states in the decoder trellis, on Light and Average shadowed channels. The performance of the system indicates that it would be necessary for the Canadian MSAT channel, however the gains on a K=10 dB Ricean channel, such as in the United States, are small. The algorithm for the receiver can be realizable using standard signal processing hardware, which would cut the cost and development time of a practical realization of the system.

VII. REFERENCE

- [1] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals", *IEEE Transactions on Information Theory*, Vol. IT-28, No. 1 January 1982, pp. 56-66.
- [2] G. Ungerboeck, "Trellis Coded Modulation with Redundant Signal Sets Part I: Introduction", *IEEE Communications Magazine*, Vol. 25, No. 2, February 1987, pp. 5-11.
- [3] G. Ungerboeck, "Trellis Coded Modulation with Redundant Signal Sets Part II: State of the Art", *IEEE Communications Magazine*, Vol. 25, No. 2, February 1987, pp. 12-21.
- [4] C. Loo, "A Statistical Model for a Land Mobile Satellite Link", *IEEE Transactions on Vehicular Technology*, Vol. VT-34, No. 3, August 1985, pp. 122-127.
- [5] A. Lee and P.J. McLane, "Convolutionally Interleaved PSK and DPSK Trellis Codes for Shadowed, Fast Fading Mobile Satellite Communication Channels", *Vehicular Technology*

Conference, IEEE, Philadelphia, Pa., June 1988.

- [6] F. Abrishamkar and E. Biglieri, "Trellis Coded Continuous Phase Modulation for Satellite-based Mobile Communications", *Globecom '88*, Hollywood Florida, IEEE, 1988.
- [7] T.A. Schonhoff, "Symbol Error Probabilities for M-ary CPFSK, Coherent and Non-coherent Detection", *IEEE Transactions on Communications*, Vol. COM-24, No. 6, pp. 644-652, June 1976.
- [8] V.K. Bhargava, D. Huccon, R. Matyas and P. Nuspl, *Digital Communications by Satellite*, Wiley Interscience, New York, 1981.
- [9] P.J. McLane, P.H. Wittke, and P. Ho, "PSK and DPSK Trellis Codes for Fast Fading Shadowed Mobile Satellite Communication Channels", *IEEE Transactions on Communications*, Vol. 36 No. 11 Nov. 1988, pp. 1242-1246.
- [10] P.J. McLane, P.H. Wittke, P.Ho and C. Loo, "Performance of Binary Multi-h Modulation for Fast Fading Shadowed Mobile Satellite Communication Channels", *IEEE Transactions on Communications*, IEEE VTC-87, June 1987 Tampa Florida: IEEE, 1987.
- [11] D. Divsalar and M.K. Simon, "The Design of Trellis Coded MPSK for Fading Channels: Performance Criteria", *IEEE Transactions on Communications*, Vol. 36 No. 8, Sept. 1988, pp. 1004-1012.
- [12] D. Divsalar and M.K. Simon, "Trellis Coded Modulation for 4800-9600 bits/s Transmission over a Fading Mobile Satellite Channel", *IEEE Journal on Selected Areas of Communications*, Vol. SAC-5, No. 2, Feb. 1987 pp. 162-175.
- [13] J. Anderson, T. Aulin and C.E. Sundberg, *Digital Phase Modulation*, Plenum Press: New York and London, 1986.
- [14] J. Thorpe and P.J. McLane "A hybrid/data Viterbi demodulator for Encoded CPFSK Modulations", *IEEE Transactions on Communications*, Vol. Comm-33, No. 1, June 1985, pp. 535-542.
- [15] W.P. Osborne and M.B. Luntz, "Coherent and Non-coherent Detection of CPFSK", *IEEE Transactions on Communications* Vol. Comm-22, 1974, pp. 1023-1036.

Light Shadowing	Bandwidth	Int. deg.	Fade Margin at 10^{-3} (dB)	Gain (dB)	Fade Margin at 10^{-4} (dB)	Gain (dB)
Non-interleaved 16 State	.01	-	10.4	N/A	18.1	N/A
(13,4) 8 state	.01	10	11.3	-0.9	15.3	2.8
	.01	18	7.4	3.0	10.5	7.5
(23,14) 16 state	.01	9	12.4	-2.0	17.7	0.4
	.01	10	11.4	-1.0	15.6	2.5
	.01	18	6.2	4.2	8.1	10.0
Non-interleaved 16 State	.05	-	8.8	N/A	15.5	N/A
(13,4) 8 State	.05	9	3.0	3.8	3.7	9.8
	.05	18	3.3	3.3	3.7	11.8
(23,14) 16 State	.05	9	3.3	3.3	6.0	9.5
	.05	18	3.4	5.4	3.5	12.0
Non-interleaved 16 State	.1	-	8.0	N/A	12.7	N/A
(13,4) 8 State	.1	9	4.0	4.0	4.3	8.4
	.1	18	3.4	4.6	3.8	8.9
(23,14) 16 State	.1	9	4.2	3.8	4.2	8.5
	.1	18	3.4	4.6	3.4	9.3

Table 2: Summary of Results on Light Shadowed Channel

Average Shadowing	Bandwidth	Inc. deg	Fade Margin at 10^{-3} (dB)	Gain (dB)	Fade Margin at 10^{-4} (dB)	Gain (dB)
Non interleaved 16 State	.01	-	12.7	N/A	19.4	N/A
(13,4) 8 State	.01	9	13.5	-0.8	18.5	0.9
	.01	18	8.7	4.0	11.2	8.2
(23,14) 16 State	.01	9	13.5	-0.8	17.6	1.8
	.01	18	7.5	5.2	9.6	9.8
Non-interleaved 16 State	.05	-	11.1	N/A	17.2	N/A
(13,4) 8 State	.05	9	5.9	5.2	7.0	10.2
	.05	18	4.2	6.9	4.6	12.6
(23,14) 16 State	.05	9	6.1	5.0	7.4	9.8
	.05	18	4.0	7.1	4.4	12.8

Table 3: Summary of Results on Average Shadowed Channel

Rician K = 10 dB	Bandwidth	Inc. deg	Fade Margin at 10^{-3} (dB)	Gain (dB)	Fade Margin at 10^{-4} (dB)	Gain (dB)
Non interleaved 16 State	.01	-	2.4	N/A	4.7	N/A
(23,14) 16 State	.01	9	5.7	-3.3	6.8	-2.1
	.01	18	3.2	-0.9	3.6	1.2
Non-interleaved 16 State	.05	-	1.9	N/A	3.6	N/A
(23,14) 16 State	.05	9	3.1	-1.2	3.2	0.6
	.05	18	2.2	-0.3	1.7	1.8

Table 4: Summary of Results on K = 10 dB Rician Channel

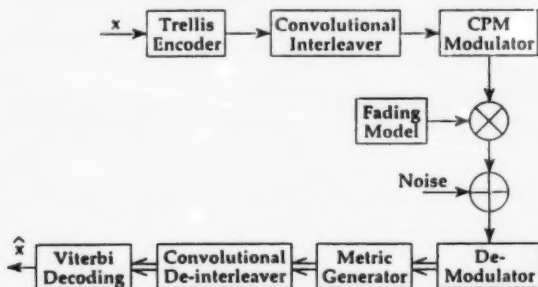


Figure 1: Block Diagram of System Model

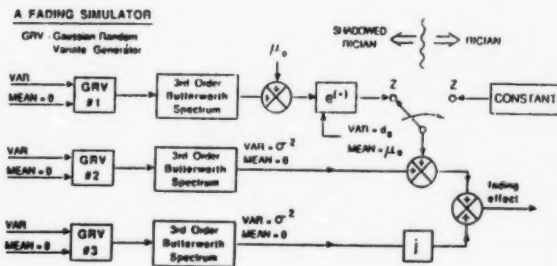


Figure 2: Block Diagram of Fading Simulator

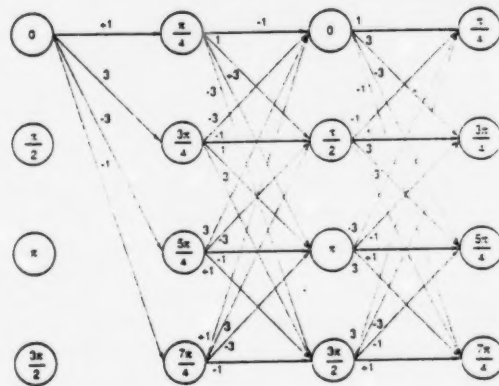


Figure 3: Modulator Trellis of Uncoded $h = \frac{1}{4}$ 4-level CPMK

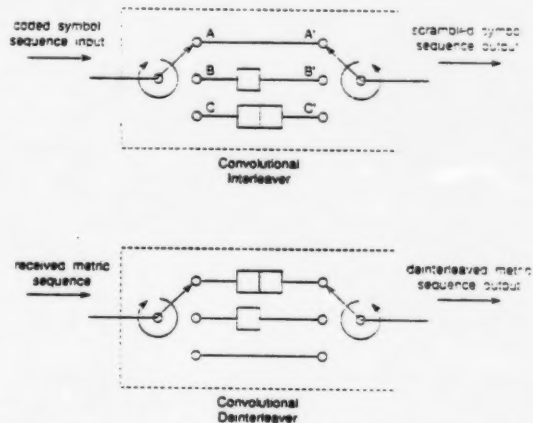


Figure 4: Example of Interleaver/De-interleaver Structure

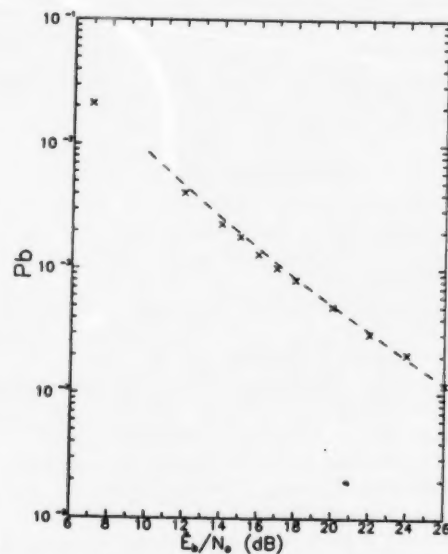


Figure 5: Comparison between Simulations and Theory for Non-interleaved Slow Fading Case

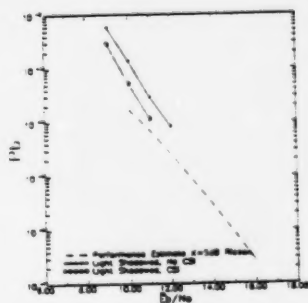


Figure 6: Performance Estimate and Simulations for Ideal Inter-leaving Case

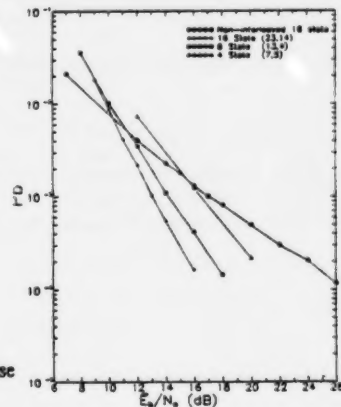


Figure 7: Results of the Interleaved System with 4, 8 and 16 state Encoders

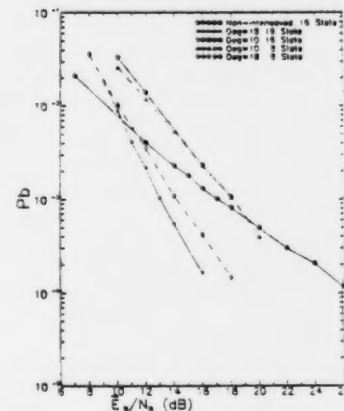


Figure 8: Results for Light Shadowing Slow Fading ($BW = 0.01$) Case

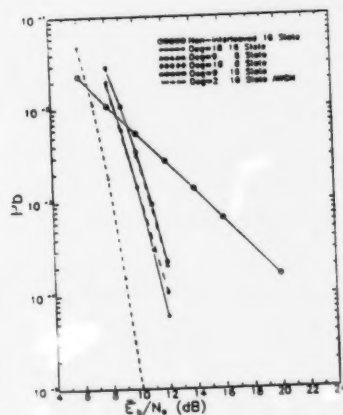


Figure 9: Results for Light Shadowing Fast Fading ($BW = 0.05$) Case

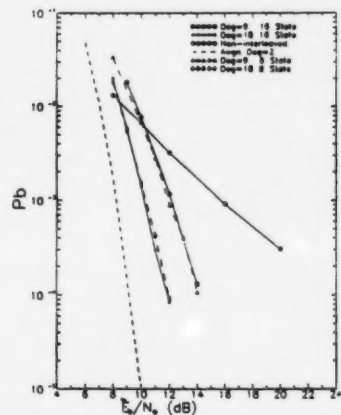


Figure 10: Results for Light Shadowing Fast Fading ($BW = 0.1$) Case

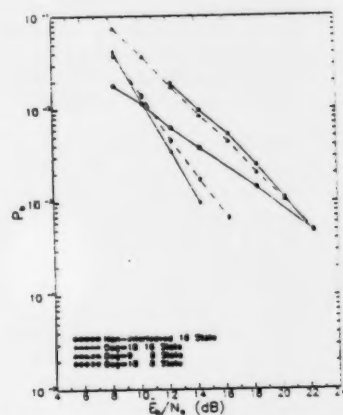


Figure 11: Results for Average Shadowing Slow Fading ($BW = 0.01$) Trials

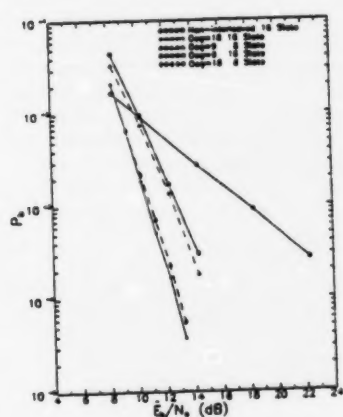


Figure 12: Results for Average Shadowing Fast Fading ($BW = 0.05$) Trials

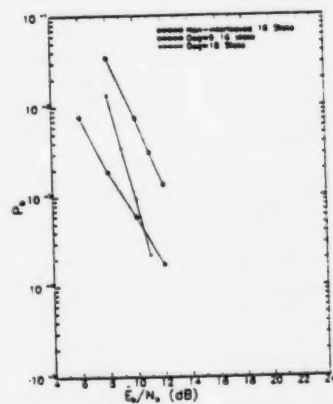


Figure 13: Results for Slow Fading ($BW = 0.01$) $K = 10dB$ Rician Trial

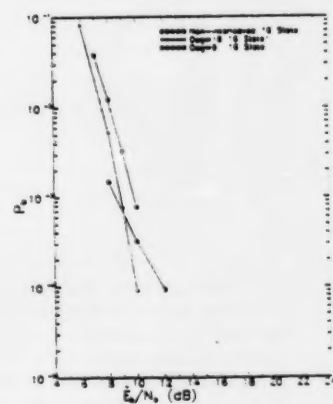


Figure 14: Results for Fast Fading ($BW = 0.05$) $K = 10dB$ Rician Trial

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

Wide Area Vehicular Communications

A.I. Spolsky, Microtel Pacific Research Ltd., Burnaby, BC
M.O. Ali, South West Bell, Kansas City, Missouri
V.C.M. Leung, University of British Columbia, Vancouver, BC

INTRODUCTION

There has been considerable activity recently in the US, Canada, Europe, Australia and Japan towards the planning and implementation of experimental and commercial mobile satellite communication systems (MSS) to complement traditional wireline switched telephone networks and terrestrial-based mobile services. The advantages of providing mobile communication services via satellite have been discussed sufficiently in related literature, one of them being the possibility of extending multiple services to a variety of geographically dispersed users with varying application needs, all through the same network.

MSS networks planned for introduction in the early 1990's are invariably expected to be of the multi-service type, support wide-area Mobile Telephone Service (MTS), Mobile Radio Service (MRS), and Mobile Data Services (MDS). MTS and MRS will be voice communication services, with MTS subscribers allowed direct access to and from the public switched telephone network (PSTN) in much the same way as possible with terrestrial-based cellular mobile services. MRS, on the other hand, will be a private radio service tailored to closed user groups, and will not provide direct access to the PSTN. MDS is intended to extend data communication capabilities between a customer's facility and mobile or portable terminals belonging to the same private network. A variety of data communication applications are planned to be supported including, but not limited to: interactive digital messaging (including vehicle dispatch), paging, vehicle location, data acquisition and control, data collection, data broadcast, and bulk data transfer (data, image, text).

The integration of multiple services, each with its own distinct characteristics and processing requirements, poses special problems to the design of the network signalling system. One the one end, there are voice-based services which may best be served through circuit-switched connections under Demand-Assigned Multiple Access (DAMA) control; on the other, are packet-oriented data communication applications characterized by short bursts of transmission occupying no more than a few hundred milliseconds of channel time. The background processing requirements and channel set-up times inherent in DAMA control renders this scheme unsuitable for serving the latter applications since the total pre-transmission delay may very well be an order of magnitude or more longer than the transmitted packet.

The design of a network signalling system for MSS must take into consideration the differences in processing requirements and traffic characteristics of the various applications under the three service categories, and the transmission constraints imposed by the severe propagation environment in which MSS Systems will operate. Key goals in the design of the signalling system must include:

- minimizing signalling delays,
- minimizing blocking of user requests (i.e. from collisions on the signalling channel),
- minimizing the number of signalling channels required in the network,
- ensuring signalling channel stability even at high traffic rates, and
- minimizing processing complexity at the terminal end.

This paper will assess the alternative signalling network designs in view of the planned multi-service mobile satellite systems and the aforementioned design goals, and suggest an approach to an efficient signalling system implementation.

MSS NETWORK CONCEPTS

A proposed network layout for first generation MSS Systems is illustrated in Figure 1. The network will consist of a satellite, a Network Control Center (NCC), a number of fixed Gateway and Base Stations, Data Hubs, and mobile user terminals for the different services. The fixed earth stations will operate at frequencies in the Ku-band, and the mobile terminals in the L-band around 1.6 GHz, with the MSS satellite providing the necessary translation between the two frequency bands. While satellites with highly eccentric orbits have been proposed to give better coverage to polar latitudes (Norbury (1)), planning efforts in North America have concentrated on satellites in geosynchronous orbits (Davies et al (2), Pattan (3)).

In the proposed architecture, the NCC will serve as the nerve center of the network, centralizing such functions as network surveillance and control, network performance monitoring, system administration, call processing and the management of satellite resources, and service billing. The NCC will communicate with the other ground segment elements via dedicated signalling links, and with the orbiting MSS spacecraft via a Satellite Control Centre (SCC).

The Gateway and Base Stations will be used for processing calls within the MTS and MRS service categories, respectively. Both types of stations will be under the control of the NCC for voice channel setup and takedown, station equipment configuration, station diagnostics, and other network control functions. The Gateway Stations will interface directly to the PSTN, and Base Stations will connect to private organizations' dispatch consoles.

Data Hubs will support applications within the MDS service category. These network elements will likely be shared among many private data networks, with data links being established between the Hub station and individual organizations. Where necessary, access to public data networks will be provided through these stations.

MSS NETWORK SIGNALLING REQUIREMENTS

The signalling network in MSS systems is defined to mean that subset of the communications network which carries non-billable signalling messages between the network elements. In the proposed architecture of first generation MSS systems, the signalling network will have two components: The L/Ku-band signalling subsystem, and the Ku-band signalling and control subsystem. Individual characteristics and requirements of these signalling subsystems are discussed below.

L/Ku-band Signalling Subsystem

This subsystem will carry signalling messages between mobile user terminals and the fixed earth stations. Traffic through this subsystem may include the following message exchanges:

- terminal log-in and log-out registrations
- communication channel requests
- communication channel assignments
- terminal control commands
- frequency and timing offset broadcasts
- acknowledgement messages

With the centralized nature of first generation MSS systems, it is expected that the majority of signalling message exchanges in this subsystem will be between mobile terminals and the NCC. Some of the more important factors and considerations to be taken into account in the design of this signalling subsystem are discussed below.

- a. Length of Signalling Messages An important characteristic of the L/Ku-band signalling subsystem is that the transmitted messages are expected to be relatively short and generally requiring fast responses. Excluding message headers, error control coding and other overhead, inbound signalling messages in this subsystem may be constructed using the following three information fields:

- message_type
- message
- terminal_ID

For all conceivable types of messages over this signalling link, a 4-bit allocation should be sufficient for the message_type field. The length of the message field will vary with the message_type. In some cases, this field will have zero bits. The worst case will occur when digits are being passed from an MTS terminal for dialing out into the PSTN. For this, an allocation of seven bytes for the message field can be made to accommodate the dialing out of up to 14 digits from MTS terminals (BCD coding assumed, conservatively). Unique terminal_ID numbers for user populations in excess of 6 million subscribers can be accommodated within a 24-bit allocation.

Under these assumptions, the longest raw signalling message can therefore be estimated to be about 84 bits in length. Assuming a rate 1/2 coding algorithm for error control, a message header of 2 bytes for bit synchronization, and a unique word of 15 bits for frame synchronization, signalling messages in this subsystem will then vary between about 87 bits and 200 bits in length. At a planned transmission rate of 2400 bits/sec over the signalling channels, this translates to packet sizes ranging between about 36 msec and 83 msec in duration, with the majority of them falling within the lower end of the scale.

- b. Number of Signalling Transactions The total number of signalling message exchanges required to setup a network connection between a mobile terminal and a fixed earth station will have an impact on the overall capacity of the signalling system. The more the number of signalling messages, the longer the channel time needed to establish each connection, and hence the fewer the number of mobile terminals that can be supported on one signalling channel. A design goal must therefore be to minimize the number of signalling message transactions required to setup a network connection.

Further, the cumulative overhead introduced in every call setup due to network signalling (t_s) will have a bearing on the real efficiency (e_p) of the signalling protocol, as illustrated by the relationship:

$$e_p = 1 - t_s / (t_s + t_c)$$

where t_c is the time needed to transmit "billable" information on the assigned communication channel. For voice calls, t_c would be the voice channel occupancy time. Note that the higher the ratio (t_s / t_c), the lower the protocol efficiency. With packet lengths in some of the MDS applications expected to be very short (in the order of 30 bytes), it is essential that the total signalling overhead be kept significantly smaller.

- c. Channel Utilization Efficiency With the advent of faster computers and affordable associated memory devices, the overall subscriber capacity of MSS networks will ultimately be limited by the capacity of the signalling channels. New subscribers can be added onto the network only for as long as there is capacity on the inbound signalling channels to provide a specified grade of service to the user population.

Although it can be argued that more and more channels can be used for signalling purposes in response to increases in customer base, it should be pointed out that the signalling channels are by definition non-revenue-generating channels. A key design goal must therefore be to make the most efficient use of the signalling channels so that a minimum number of them will be required to support a large subscriber base. This way, more system channels can be left for assignment to revenue-generating communications.

Other factors which will contribute to the overall capacity of the signalling network include transmission (data link) protocols, choice of multiple access schemes for the inbound signalling channels, and the manner in which the different services are integrated for signalling purposes.

Ku-band Signalling Subsystem

In the proposed architecture of first generation MSS systems, the Ku-band signalling subsystem will carry signalling messages between the NCC and the Gateways, Base Stations and Data Hubs. Traffic over this signalling subsystem is expected to include the following:

- communication channel assignments
- incoming call (from PSTN) notifications
- communication channel requests
- communication channel release messages
- call progress messages
- call records and billing data
- station alarm messages to the NCC
- station diagnostics commands
- station reconfiguration commands
- acknowledgement messages

As can be observed from the above list, transactions in the Ku-band signalling subsystem will consist of both short and very long messages, in a mixture of ones which require fast responses and those which can be processed in background. The difference between a typical short message (e.g. communication channel assignment message) and a long message such as a call record could conceivably be as large as an order of magnitude or so in duration.

Important design goals for this signalling subsystem are: to ensure that real-time messages are given fast access over the signalling links; to ensure high integrity of data transmission over the signalling links; and to minimize the overall number of signalling channels required in the network.

SIGNALLING SYSTEM DESIGN ALTERNATIVES

The design of a network signalling system can be specified in terms of call setup algorithms, multiple access techniques, modulation schemes, and signalling rate. For first generation MSS networks, it is assumed that 5 kHz channels will be used, and the DMSK will be the modulation scheme for 2400 bits/sec data transmission within these narrow bandwidth channels. No discussion of alternative modulation schemes or signalling rates will be carried out

in this paper. The remaining signalling system parameters have not been as firmly up and will be used to discuss alternative designs for the two signalling subsystems proposed in first generation MSS networks.

1. Call Setup Algorithms

There are some important differences in characteristics of "billable" traffic between voice and data calls which necessitate the use of different algorithms for call setup. These differences are elaborated under the following sub-headings.

- a. Voice Calls MTS and MRS calls are characterized by long channel occupancy times (averages of 3 minutes, and 20 seconds, respectively), and are generally suited to processing through a DAMA control scheme. For these, the extra channel setup delay due to DAMA processing is considered inconsequential because of the long voice channel holding times.

A typical (abbreviated) call setup signalling sequence for an MTS call is outlined below. For simplicity, a terminal-originated call into the PSTN network is illustrated:

- user terminal sends a channel request message to the DAMA controller with information on the digits to be dialed
- DAMA controller determines the network resources to be used for the call and responds to the requesting terminal with a voice channel assignment message
- DAMA controller sends to the appropriate earth station (via the Ku-band signalling subsystem) a channel assignment message with the digits to be dialed out and the identity of the originating terminal
- the requesting terminal moves to the assigned voice channel, and using signalling message exchanges on the voice channel, the processing earth station verifies the identity of the terminal on the channel
- the processing earth station then dials out the called number and allows direct communication between the terminal and the PSTN line appearance.

In a recent study by the authors (Leung et al (4)), some basic differences in the applications, network connectivity, and traffic characteristics of MTS and MRS services were identified. As a result, a slightly different call setup procedure for MRS service was recommended, incorporating call queuing, batch processing, and pipelined messaging. A complete presentation of the differences and discussion of recommendations can be found in reference (4).

Note that in both call setup procedures, there is attempt to shift most of the signalling exchanges from the less robust L-/Ku-band signalling channel to the assigned voice channel. This is because in-band signalling on the latter channel will be free of the many limitations of the contention signalling channel, and hence more efficient.

- b. **MDS Call Setup** Information packets transmitted from MDS terminals are expected to be in the order of 30 to 250 bytes, depending on application. At a transmission rate of 2400 bits/sec, this amounts to channel occupancy times in the range 100 msec to about 833 msec. Considering that the minimum delay incurred by a message, excluding message transmission time, is at least twice the satellite channel propagation time, and that a typical terminal may require about 200 msec to switch between channels, data calls cannot be efficiently setup using the same DAMA schemes as proposed for voice calls.

A number of proposals have been made to try to integrate the data services into the multi-service architecture of MSS Systems. In a recent investigation by the authors (Ali et al (5)), it was concluded that the most appropriate way of handling MDS services would involve the uncoupling of signalling tasks, the grouping of applications into sets of bearer services with common characteristics, and controlling the real-time allocation of channel time to MDS users from the serving Data Hubs. Under this proposal, data applications were classified under inbound simplex (ISX), outbound simplex (OSX), and full duplex (FDX) bearer services, and individual call setup procedures were then defined to cater to the unique characteristics of each bearer service. These procedures are closely tied in with the access protocols for data packets, and will be discussed under the following heading.

2. Multiple Access Techniques

Multiple access protocols have been widely used in satellite communication systems for conflict resolution among users desiring transmission capacity on the often few available communication channels. Techniques such as FDMA, CDMA, TDMA and ALOHA-based protocols have all been used to satisfy different system requirements and traffic patterns. Contention protocols (eg. random and slotted ALOHA, random access TDMA) are usually the least complex to implement, impose the least delay, but provide very lower throughputs (no better than a theoretical maximum of 36%). Fixed assignment and reservation protocols are somewhat more complex, have longer delays, but yield higher throughputs in excess of 60%. Further, it has been shown that some traffic characteristics do favour one class of protocols more than others (Lam (6)), as summarized in Table 1.

An examination of the traffic expected on MSS signalling subsystems shows that a mixture of traffic patterns will coexist, and that variations will occur between the two subsystems and also within the

Table 1: Traffic Models and Multiple Access Protocols

Traffic Model	Favoured Multiple Access Protocols
Non-bursty	Fixed assignment protocols (FDMA, TDMA)
Bursty, short messages	Contention protocols
Bursty, long messages, large number of users	Reservation protocols with contention reservation channel
Bursty, long messages, small number of users	Reservation protocols with fixed TDMA reservation channel

service categories. This suggests that a more appropriate way of specifying multiple access protocols for the network is to consider the requirements of each signalling subsystem separately.

- a. **L-/Ku-band Signalling Subsystem** As indicated earlier, signalling messages in the MTS and MRS services are expected to be short in duration, variable in length, and bursty in occurrence. This type of traffic is best suited to pure contention protocols. Pure ALOHA and random access TDMA are recommended. Note that slotted ALOHA may not be the most efficient contention access technique in this application since it requires using slot duration based on the maximum packet size.

For the MDS service, there are two separate signalling requirements: (i) signalling between MDS terminals and the NCC (for system registration, etc.), and (ii) signalling between MDS terminals and the Data Hubs. The former requirement is similar in characteristics to signalling encountered with voice terminals, and may therefore be handled efficiently via a contention protocol. Access schemes for the latter requirement have been discussed in detail in reference (5).

- b. **Ku-band Signalling Subsystem** Signalling messages in this subsystem are widely varied in size. Messages exchanged during call setup and take-down are expected to be short, similar in size to those exchanged between the NCC and mobile terminals through the L-/Ku-band signalling link. Most of these messages require immediate responses from the DAMA controller. Other messages on this link (e.g. call records), are typically considerably longer, but generally do not require real-time responses, if any. Further, signalling traffic in this subsystem will be concentrated in much fewer nodes with Gateway Stations generating the most.

To guarantee fast accesses for the potentially high concentration of real-time signalling traffic at each Gateway Station, it is recommended that some preassigned capacity (e.g. fixed frame TDMA) be allocated for this portion of Ku-band traffic. Short messages from Base Stations and Data Hubs are expected to be infrequent, and generally less sensitive to processing delays. For this traffic, a contention scheme may be employed. Long, non-realtime messages from Gateway and Base Stations may be efficiently

accommodated via a reservation protocol.

SUMMARY AND RECOMMENDATIONS

Characteristics of the MSS network signalling system have been presented, and possible implementations of its subsystems have been proposed in terms of signalling and call setup procedures, and multiple access techniques. Because of the severe L-band propagation environment, it is recommended that signalling messages on the L-/Ku-band signalling channel be kept to a minimum, and that the majority of signalling exchanges be performed in-band on the assigned voice channel and relayed back to the NCC via the more robust Ku-band signalling subsystem.

Call setup procedures and multiple access techniques have been defined taking into consideration the differences in characteristics of the services to be supported. It was shown that because of these differences, a number of signalling sub-networks will be needed to efficiently address the service requirements. Further work is needed to determine optimum packet sizes for the signalling messages. This will give more insight into the traffic carrying capacity of the signalling subsystems and provide estimates for the total number of signalling channels required in a network.

REFERENCES

- (1) Norbury, J.R., 1986, "Land Mobile Satellite Systems Using Highly Eccentric Orbits", Proceedings of an ESA Workshop on Land Mobile Services by Satellite, 132-136.
- (2) Davies, N.G., Shoamanesh, A., and Leung, V.C.M., 1988, "Development of the Network Architecture of the Canadian MSAT System", Proceedings of the JPL Mobile Satellite Conference, Pasadena, USA.
- (3) Pattan, B., 1987, "The Advent of Land Mobile Satellite Service Systems", IEEE Transactions on Aerospace and Electronics Systems, Volume AES-23, No. 5, 691-703.
- (4) Leung, V., Spolsky, A. Ali, M., and Acres, G., 1988, "Demand Assignment for Mobile Radio Networks over Satellite", Proceedings of the IEEE 38th Vehicular Technology Conference, Philadelphia, USA.
- (5) Ali, M.O., Leung, V.C.M., and Spolsky, A.I., 1988, "A Framework for Implementing Data Services in Multi-Service Mobile Satellite Systems", Proceedings of the JPL Mobile Satellite Conference, Pasadena, USA.
- (6) Lam, S.S., 1979, "Satellite Packet Communication - Multiple Access Protocols and Performance", IEEE Transactions on Communications, Volume COM-27, No. 10, 1456-1466.

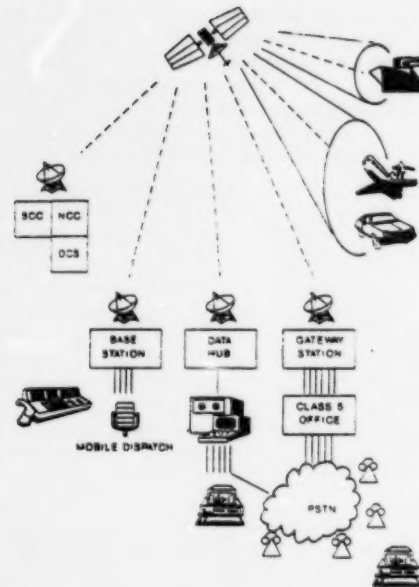
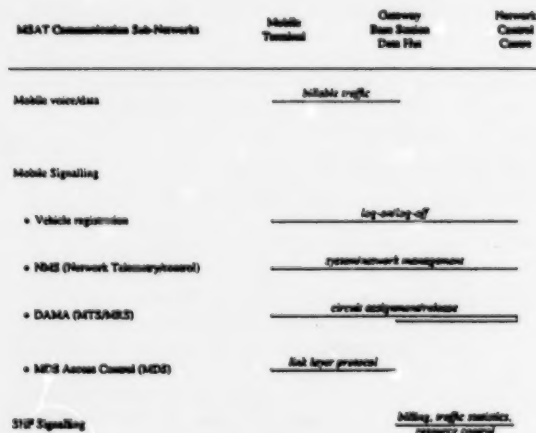


FIGURE 1: MSS Network Layout



MSAT NETWORK CONNECTIVITY

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

A HARD REAL-TIME OPERATING SYSTEM FOR AIRBORNE
SCIENTIFIC EXPERIMENTS

J.E. Jordan
Flight Research Laboratory,
National Research Council of Canada,
Montreal Rd., Ottawa, Canada K1A 0R6

This paper describes a hard real-time, multi-task operating system developed at the Flight Research Laboratory of the National Research Council of Canada (NRC) for airborne scientific experiments in oceanography, geophysics and navigation. This system has been used extensively for a number of years in a variety of experiments, primarily in experimental magnetic surveys ranging from the high Canadian Arctic to the Caribbean.

A number of aspects of this system are described with particular emphasis on task organization and scheduling, software development methods and a special operational mode, called "ground refly", used in exercising and testing flight software. One of the main requirements of this system was the ability to process sampled data in real time within the time frame defined by the sampling clock period. Failure to perform this processing by the next clock interrupt could result in serious errors which could jeopardize the objectives of experiments.

Software development for this system is an important consideration because of the nature of the system. A semi-mechanized configuration control system is described. Since debugging of software in flight is undesirable due to the cost and rigours of flight operations, a special "ground refly" mode of operation was developed to simulate the flight environment by replaying recorded flight tapes. This mode, which exercises most flight software from low-level device interrupt handlers to the application tasks, is described.

Introduction

This paper begins by describing a simple hard real-time, multi-task operating system used to control a computer system in an aircraft "flying laboratory"[1], used in a variety of scientific experiments at NRC. Much of this work has involved experimental magnetic surveys for ASW and geophysical applications where magnetic sensors are used to measure the magnetic field of the earth and its spatial gradients which must be referenced to the position of the aircraft as measured from a variety of navigational devices. Subsequent sections deal with aspects of the system which are particularly relevant to operating system development and emphasize the hard real-time nature of this system.

The first aspect considered is the task organization and scheduling employed in the system. Subsequently, software development aspects of the system and a special operational mode called "ground refly" are described. General details of the system and its operation are available elsewhere [2].

The operating system for this application was developed at NRC and has been operational over a number of years. The primary objective of the system was to process a number of periodically sampled streams of data from/to a variety of input sensors/output devices in real time. As well, it was considered essential that this processing should occur in lock-step with the input sampling, within the time interval defined by the sampling clock period. This is essentially a hard real-time system since this processing must be completed before the deadline imposed by the next periodic sampling clock interrupt. The system can also handle asynchronous external events within its framework though

no attempt has been made to use clock interval event deadlines (though this could be done in principle). The operating system provides a number of command line processing tasks which operate asynchronously on "end of command line" events and handle interactions with a project operator, pilots and navigator. A variety of sensors, signal conversion equipment, digital tape recorders and display units are handled by this system. In general, the approach used in this system has worked well for our purposes.

It is noted that some hard real-time systems are used in safety-critical applications such as flight control, traffic control, process control and factory automation requiring fault-tolerant features such as redundant hardware and "n" independently-written versions of software. In this application, failure of the system to meet computational deadlines could seriously jeopardize scientific experiments requiring costly rescheduling of flights and experiments. For instance, one computational task involves the active compensation of the outputs of the magnetic sensors in real-time, which removes the effect of the magnetic field of the aircraft in order to measure the ambient magnetic field of the earth without error. Other instances are those navigation tasks which compute aircraft position by integrating measured accelerations and velocities ("dead-reckoning"). Failure to perform these computations occasionally is not catastrophic; however, missed computation cycles lead to cumulative errors which after several minutes would become unacceptable. The positioning data computed by the system is used mainly for recording with other sensor data, though it is also displayed to the pilots and navigator. The accuracy of this positioning data is obviously of concern but is not considered safety critical since the primary navigation avionics equipment operates independent of the computer system and is redundant as required by flight safety regulations. The simple non-redundant nature of this system reflects the design philosophy that this system need not address safety-related issues.

Task Organization and Scheduling

Software in this system is structured into a number of tasks which compete for resources and which are scheduled and run under the control of the operating system. The system mainly operates in conjunction with a number of sensors and conversion equipment which are sampled periodically at various rates to create a number of data streams in and out of the processor as tabulated in Table 1. Associated with each periodic data stream is a hardware clock which initiates the sampling and which interrupts the computer generating a clock event which sets up the corresponding periodic processing task. This processing task executes a number of non-data-dependent procedures before performing a SIGNAL & WAIT system procedure to synchronize itself with the latest sensor input data (as indicated by a "data-available" flag). The task blocks or WAITs if data is not yet available. When the sampling device completes its operation, it generates an interrupt to the processor which sets the "data-available" flag. This event performs a SIGNAL which unblocks the waiting task. The task can then execute a number of procedures to process the input data as required. However, processing of sensor data is constrained to be completed before the next sampling clock interrupt when the task would normally be restarted. If not, the task is aborted and restarted generating an error message to the operator. The operation of the scheduler is illustrated in Table 2.

The scheduler used in this system determines which task should be run next based on priorities of tasks. The scheduler operates under a preemptive strategy but incorporates logic to recycle tasks which have not completed processing within their sample clock period. For periodic tasks, priorities are assigned with the highest priority being given to the highest rate task and so on. This, in effect, gives priority to tasks whose deadlines are closest (i.e. next clock tick of highest rate task). This allows hard real-time scheduling using a simple, well-known approach. This approach to scheduling is similar to that used in the SIFT [3] system for flight control. The periodic tasks are constrained to be "simply periodic" (iteration rate of a given task is an integral multiple of the next smaller rate). Wensley et al [3] note that this arrangement guarantees that all tasks will complete within their designated periods provided the workload doesn't exceed the capacity of the processor.

Tasks which must be run in response to asynchronous external events can also be handled using this scheduler though care must be taken in assigning priorities. In our system, command line processing tasks are scheduled upon interactive terminal command line inputs and are assigned a priority lower than the real-time periodic tasks (except for the 0.1 Hz, ten second period task). This provides adequate response time and does not interfere with the hard real-time periodic tasks. It is noted that excessive processing at one of the periodic task levels can lock out the command processor so an Urgent Command Processor operating at the maximum system priority has been implemented to handle this eventuality. An "urgent command" is manually invoked by prefixing a given command with 'U'. The task scheduler computes the elapsed processing time of each task by reading a high-resolution counter timer upon each context switch. This information may be displayed by the operator to aid in adjusting the workload or deciding whether to add more procedures at a given priority level. The priorities of tasks in this system are shown in Table 3.

The tasks themselves are each subdivided into a number of procedures which are run consecutively. This is formalized in the system by using a procedure scheduler which operates on chains (tables) of procedures which are set up for each task. Three (or more) separate chains are defined for each task including a non-data-dependent chain (NDC), a chain for sampling device number 1 (D1C), a chain for sampling device number 2 (D2C) and more sampling device chains as required. The first chain of procedures is executed when the task is started and each device chain is executed when its associated device data becomes available (the first procedure in each device chain issues a "data-available" flag WAIT primitive which synchronizes the task by blocking until a corresponding SIGNAL event occurs). The chains themselves are implemented as tables of procedure addresses terminated by a zero entry. The procedure scheduler is a reentrant routine which operates using register-based pointers to follow the chains from head pointers maintained in each task control block. Procedures on these chains may be turned on/off by an unused bit in the address entry on the chain. This allows some flexibility in determining which procedures are executed for any given circumstance. The system operator can turn procedures on or off under command control as required. A typical series of procedures for the 8 Hz task is illustrated in Table 4.

This approach has worked out well for our purposes, generally speaking. There have been instances of "bursty" tasks (having low average processing times but having worst case processing times exceeding the task period) not suitable for the hard real-time scheduling approach. These tasks were handled by buffering the periodic data and processing it using an asynchronous task. No attempt has been made to incorporate deadlines for asynchronous tasks, though in principle, it should be straightforward to do this in the scheduler using an interval timer.

Software Development

This system was largely written in Assembler though many of the application tasks are written in FORTRAN. At the time development was started, other language choices were not available or viable. It is now quite clear that there are much better choices such as C. The present system incorporates approximately 50,000 lines of source code (approximately 20 person-years of work). Though this is not a huge system, management of this amount of code is non-trivial.

As a result, a number of measures were taken to track and manage all flight software in the system. A separate ground-based computer system is used for software development since the aircraft may be in use and since it is not very convenient as a software development environment for several programmers. Baseline files of all source code, libraries and link procedures are maintained on a master account read-only accessible to all developers on the ground-based system. Individual users may copy files to their own accounts, changing them and re-linking to test or verify their code. Updates to any given file may be submitted to the group manager for inclusion in the next baseline (or revision) of the software. This approach is rather ad hoc and does not address simultaneous updates by more than one user or formally mechanize configuration control. However, for our small development group, the system has worked well. Backups of files are done after each baseline as well as on a periodic basis so it should be possible to generate any given revision of software in time if required.

Most of the application software for the system is written in FORTRAN by the scientists involved in the experiments. Much of this software is developed in an unstructured manner on a mainframe time-sharing system. This software is developed and tested off-line using conventional sequential programming methods. When the software is moved to the aircraft system, it is hoped that concurrency aspects are not a major problem within the tasking framework. However, it is still necessary to recompile the code using a different compiler and to verify that it will run in hard real-time. Most application software is written as procedures to be run at one of the task levels previously described. Once the software has been configured into the real-time flight system, it is placed under change management control as previously noted.

"Ground Refly" Mode of Operation

The "ground refly" mode of operation has proven to be a useful means of testing and exercising flight software. This is quite important given the high cost and physical rigours of flight operations. This mode of operation is basically a simulation in that recorded flight tapes are replayed to re-create the flight environment on the ground. A specially linked version of the system is set up to achieve this. A high frequency clock is used to derive the timing for the sampling clocks and sensor devices normally used in flight. Data which would normally be input from the various sensors is derived from the replayed tape through special input double buffers. Interrupts from the special clock are counted down and simulated interrupt instructions for all sample clock and sampled devices are issued by this interrupt handler.

These interrupts are handled by the sample clock and sampling device drivers as in the normal system though some modifications are incorporated to get the device data from the magnetic tape buffers. Most of the code used is the same though some modifications have been made for the refly system by means of conditional assembly/compilation. The refly system is also linked by a separate link procedure which incorporates the modified object modules required for refly. The refly simulation can be run at multiples or sub-multiples of real time by changing the count-down factors used by the special clock interrupt handler.

In operation, the refly system is much like the flight system. Raw and processed data generated by the system are recorded on magnetic tape as in flight and this feature can also be used for reprocessing data using new software or algorithms as required. The refly system takes its initial time from the first record found on the replay tape which may be pre-positioned as desired. Outputs to displays or chart recorders occur under refly just as they would in flight. The system can be paused and restarted which is a convenient feature which is not possible in real-time. The important aspect of this refly system is that it operates in real-time and exercises most of the flight software from low-level interrupt handlers to the application tasks making for a realistic simulation. Since the floating-point hardware on the laboratory system is different from that of the aircraft system, some timing differences are apparent. The refly system can also be run on the aircraft computer system when it is in the hangar allowing for a more realistic simulation if timing is of major concern.

Conclusion

This system has been in operation for several years. The scheduling of periodic tasks and the concept of flexible chains of task procedures has worked out well in practice. This system was designed to be relatively small and therefore contains minimal support for filing (contiguous files only), command interpretation (no command files) and fault handling (detection rather than correction). Device sampling is done under the control of a hardware clock which has turned out to be a relatively rigid arrangement. Ideally, clocking could be done under program control if a computer with priority interrupts were available. Future plans include moving to a more modern processor with improved performance and serviceability, using a front-end processor as a means of streamlining and isolating interrupt handling from the main processor and using a graphical user's interface to allow those not intimately involved with the system to use it.

Acknowledgements

The author is grateful to Mr. C.D. Hardwick for his support and leadership of this project. He would also like to thank all those involved in the implementation of the system. The facilities described in this presentation support joint research programs funded by the Department of National Defence, the Department of Energy, Mines and Resources and other industrial partners.

References

- [1] C.D. Hardwick, "Research Capabilities of the NAE Convair 580", DME/NAE QB-76(4), National Research Council of Canada, 1976
- [2] J.E. Jordan, ORQS Operator's Manual, LTR-FR-97, National Research Council of Canada, Sept. 86
- [3] J.H. Wensley, L. Lampion, J. Goldberg, M. Green, K. Levitt, P.M. Melliar-Smith, R.E. Shostak and C.B. Weinstock, "SIFT: Design and Analysis of a Fault-Tolerant Computer for Aircraft Control", Proceedings of the IEEE, vol. 66, no. 10, Oct. 1978, pp. 1240-1255.

Table 1: Devices:

Device	Rate	Description
Analog/Digital Converter #1	8 Hz	32 chans, 14 bit, +/-10 volts
Analog/Digital Converter #2	8 Hz	8 chans, 16 bits, +/-10 volts
Digital/Analog Converter #1	8 Hz	4 chans, 12 bits, +/-10 volts
Frequency/Digital Converter	8 Hz	3 chans, 24 bits
Discrete Input/Output	8 Hz	32 bits in/32 bits out
8 Channel Recorder	var	8 channels, digital
Synchro/Digital Converter	2 Hz	3 chans, 16 bits, +/-180 degs
Digital/Synchro Converter	2 Hz	3 chans, 16 bits, +/-180 degs
Doppler Radar (pulse/digital)	2 Hz	3 chans, 16 bits
Static Pressure (counter)	2 Hz	1 chan, 16 bits
Magnetometer Counters	0.5 Hz	3 chans, 24 bits
VLF NAV Receivers	0.1 Hz	15 chans
LORAN_C NAV Receivers	0.1 Hz	various
GPS NAV Receivers		asynch
INS NAV	0.1 Hz	various outputs
Alphanumeric Terminals		asynch ASCII
Graphics Terminal		asynch RS-232 link
Magnetic Tape Recorder		0.25 Hz 9-track, 800 BPI
Disc System		asynch parallel DMA interface

Table 2: Scheduler Operations

1. Task States:
 - 1.1 Dormant: Task resident in memory but inactive.
 - 1.2 Ready: Task 'Ready' to run when its fixed priority highest
 - 1.3 Running: Task 'Running' on processor
 - 1.4 Suspended: Task 'Suspended' awaiting some event.
2. Scheduler Operations:
 - 2.1 READY: Makes 'Dormant' task 'Ready'
 - 2.2 DISPATCH: Runs the highest priority 'Ready' task, possibly performing a context switch (saving the state of the interrupted 'Running' task and installing a new 'Ready' task in its place.)
 - 2.3 SUSPEND: Suspends a 'Running' or 'Ready' task to await some event.
 - 2.4 TERMINATE: Makes a task 'Dormant'. This can either be by a task-initiated supervisor call to inform the scheduler that a task has completed its computation or by a system-imposed action to shut down a 'Ready' or 'Suspended' task.
 - 2.5 EVENT: Makes a 'Suspended' task 'Ready'.

3. Periodic Clock Operations:

A periodic clock event normally initiates a 'Dormant' task with a call to READY followed by a call to DISPATCH. If the task is 'Running' or 'Ready', a special "clock period violation" message to the operator is generated and the task is restarted. Note that the task will be restarted in the 'Running' state if it is already the highest priority task in the system; otherwise, it will be restarted in the 'Ready' state.

Table 3: Tasks:

a) <u>periodic tasks</u>		b) <u>asynchronous event tasks</u>	
<u>pri</u>	<u>mnm name</u>	<u>pri</u>	<u>mnm name</u>
2	LVS8	8 Hz Task	1 UCP Urgent Command Task
3	LVSYS	System Task	
4	LVS2	2 Hz Task	6 PCP Pilot's Command Task
5	LVS05	.5 Hz Task	7 NCP Navigator's Command Task
			8 CP (Operator's) Command Task
9	LVS01	0.1 Hz Task	10 LVDBG Debugging Level Task
			11 LVBG1 Background No. 1
			12 LVBG2 Background No.2 (Dead Loop)

Note: highest priority=1

Table 4: Typical Chain of Procedures: (Task No. 2: 8 Hz)

<u>PID</u>	<u>Enabled Chain</u>		<u>Name/Purpose</u>
26	Y	Non-data-Dep.	DTOATASK: D/A Output
16	Y	" "	RMSTASK: 8 Chan Recorder Output
4	Y	" "	TFM: Freq/Digital Input
30	N	" "	HNYWL: INS Experiment
5	N	" "	MTUWRITE: Mag-Tape Recording
130	Y	" "	FQWT: Fuel/Quant. Weight Calc.
8	Y	Dev.No.1 Dep.	LVS8.T1: D1 Synchronization
40	N	" "	NIL: nil
9	Y	Dev.No.2 Dep.	LVS8.T2: D2 Synchronization
39	N	" "	OSRMSTASK: External Experiment
24	Y	" "	SWCHPP: Magnetic Compensation
31	N	" "	VGMAP: Vertical Grad. Mapping
32	N	" "	RFDATA: LSE Accumulation

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

FEASIBILITY OF USING SEGMENTATION TECHNIQUES FOR THE ENHANCEMENT OF MARINE RADAR IMAGES AS ENCOUNTERED IN ICE INFESTED WATERS

MIKE HUCULAK, AJIT NILAKANTAN,
*Cimmetry Systems, 8300 Cote de Liesse, Suite 102,
St. Laurent, Quebec, Canada H4T 1G7*

CHARLES GAUTIER
*Canarctic Shipping, 350 Sparks Street, Suite 1005,
Ottawa, Ontario, Canada K1R 7S8*

and

MAURICE AUDETTE
*Advanced Technology Division, Transport Development Center,
200 Ouest Boul. Rene Levesque, ste 601 Montreal, Qc., Canada H2Z 1X4*

Abstract — The objective of this research is to investigate the feasibility of using image processing techniques in order to enhance and segment the ice features present in marine radar images. To this end, various image enhancement and segmentation techniques have been selected for implementation. Enhancement methods based on linear and non-linear spatial filtering, frequency domain filtering, histogram treatment, thresholding, and mathematical morphology have been tested. The segmentation process was subdivided into two categories: edge-based and region-based techniques. Edge-based segmentation methods include convolution with local operators, relaxation labeling on edges and curves, and contour tracing. Region-based segmentation techniques include clustering, region splitting, and region merging.

I. Introduction

In order to navigate a ship through ice infested waters in the arctic region, the operator makes use of a synthetic aperture radar (SAR) image taken overhead from an airplane. The operator must locate the position of the ship on the SAR image by comparing it with a marine radar display. The first goal of this research is to develop a marine radar display system that will allow an experienced operator to perform this task more efficiently. To achieve this, we have studied the feasibility of using image processing to improve the quality of the marine radar display, which we will refer to as visual enhancement throughout the paper.

The second goal of this research is to automate the procedure of positioning the ship on the SAR image. To this end, we have studied methods for segmenting the marine radar image in order to identify ice floe features. These features can then be matched with similar features derived from the SAR image in order to automatically correlate the two images. This task will be referred to as automatic tracking throughout the paper.

With these two goals in mind, we summarize the utility of a number of different image processing techniques. We divide the

methods under consideration into three broad classes, pre/post-processing methods, edge-based methods and region-based methods.

II. Pre/Post-Processing Methods

In order to prepare the marine radar image for further treatment, pre-processing and post-processing algorithms can be utilized. The algorithms that have been considered include methods of histogram manipulation, thresholding, spatial and frequency domain smoothing (filtering), and mathematical morphology. The pre-processing and post-processing algorithms were developed after statistical studies were performed on the marine radar images of concern to obtain a quantitative measure of the features of interest.

The histogram of an image provides information about the number of pixels corresponding to each possible grey level. Histogram equalization [1,7] involves modifying the histogram so as to increase the contrast between areas of different intensity and enhance the features contained therein. It was found that the methods based on histogram treatment such as equalization or hyperbolization did not show good enhancement on the marine radar images. This was due to the fact that the image histograms were dominated by noise.

Thresholding [10,14,23] is often an effective way of extracting the useful features in an image from the background features or the noise. The different schemes used to select the threshold that were considered are the average threshold, Otsu's threshold and the adaptive threshold. The simple thresholding of the marine radar images did not adequately segment the image and hence is deemed inappropriate. Thresholding the gradient image did not show good results either, but may be useful when applied after a suitable segmentation process such as region splitting.

Various smoothing procedures were investigated because of the noisy nature of the images. Frequency domain smoothing [2,13] involves transforming the spatial image into the frequency domain and then filtering the result. An example of frequency domain

smoothing is applying a low-pass filter to remove noise. A spatial domain filter operates on the spatial image directly without performing any transformations. Examples of spatial filtering include the average filter, the median filter [9] and the edge preserving filter [12]. The results showed that the marine radar images are not suitable for frequency domain filtering but some spatial domain smoothing algorithms have shown encouraging results. The frequency domain methods are inadequate for this application because they are global techniques and hence cannot account for local effects. Good smoothing results were obtained using median filtering and edge preserving filtering with a window size not exceeding 5x5. The average filter had the undesirable effect of blurring the image too much and hence is not appropriate for this application.

Mathematical morphology [22] can be applied to an image in order to remove small features such as noise by grouping together and unifying similar regions. This process involves repeatedly expanding and contracting the image locally. The results of applying mathematical morphology to the radar images showed that it was able to close the ice-floe boundaries and better define them.

III. Image Segmentation

The aim of image segmentation is to divide images into several meaningful areas. With image segmentation, objects of interest can be extracted and submitted to further analysis or interpretation. Thus image segmentation represents an important early stage in image analysis and identification.

Two categories of segmentation techniques were applied to the marine radar images: techniques based on edges and techniques based on regions. The advantages of region-based image segmentation algorithms are their robustness to noise and the fact that they fully segment the image. Edge-based segmentation techniques, on the other hand, are easier to implement and lend themselves rather well to hardware implementation.

The image segmentation techniques were applied to marine radar images in order to separate and identify the existing ice floes in the image. The purpose of identifying the ice floes is to recognize them on the corresponding SAR image.

A. Segmentation Based on Edge Techniques

In general, edge-based segmentation is a two stage process. First, every image pixel is assigned a gradient value which is obtained from an edge detector. This gradient can be vectorial, where both the intensity and the direction of the edge are considered. After the first stage, the resultant image is an incomplete segmentation of the original image. The second stage involves a contour tracing algorithm in order to connect the scattered edges in a valid manner and produce closed contours determining the ice floes existing in the marine radar image.

For the first purpose of generating the gradient image, different types of local gradient operators can be used. The operators considered were: the Prewitt, Sobel, Laplacian and biLaplacian operators [3,17,18]. Due to the differential nature of the gradient operators, they tend to enhance scattered noise as well as edges. The application of the edge detecting operators to the image yielded poor results because of this fact and because the boundaries in the image are ill defined. It was found, however, that the Sobel and Laplacian operators can be used to highlight ice floe boundaries after the application of a region-based method or after a pre-processing technique such as edge preserving smoothing, as they appeared to

be more robust and noise resistant.

A relaxation labeling procedure [8,21] was also utilized after the edge detection process in order to enhance the edges and eliminate the scattered noise. By taking into account neighborhood information and using conditional probabilities (in the form of compatibility functions), relaxation is able to reduce noise and ambiguous labelings in the image. However, in order to effectively use relaxation, there are many parameters that must be fine tuned. One step that is important to the process is the initial 'guess' that is used to start the relaxation. The method used for the initial guess must be chosen after examining the effect of applying various edge detecting operators on the image. In addition, the compatibility function, which relates the effect of one label on neighboring labels must be chosen through experimentation.

The relaxation labeling for edge and curve enhancement yielded the most favorable results when applied to the original marine radar image as a method for defining and enhancing the scattered edges found by the edge detector. The results obtained after relaxation were, however, not suitable for the application of contour tracing to segment the image. After many trials, it was concluded that the edge-based segmentation techniques, even when used with edge and curve enhancement techniques, are not suitable for the segmentation of marine radar images.

The results seem to indicate that edge-based segmentation is not feasible with marine radar images and hence is not suitable for automatic tracking purposes. Edge-based techniques are, however, useful for visual enhancement purposes.

B. Region-Based Techniques

Region-based approaches do not rely on local changes in intensity to segment an image but instead attempt to group regions in the image with homogeneous properties. As a result, region-based techniques are more robust to noise and the results are very suitable for contour tracing, if desired.

There are three major methods of region segmentation that were considered: clustering, region splitting, and region merging. All of these methods involve grouping pixels with similar properties into regions. Each pixel in the image is assigned a feature vector and in the feature vector we can incorporate information such as placement, grey level, and similarity with increasing size and decreasing weight of neighborhood windows. The feature vectors are then classified using a certain measure of similarity. Similar feature vectors define the distinct regions associated with the image.

Clustering is a general, region-based scheme that groups feature vectors together under a minimum distance measure. Two different clustering algorithms were tested with the radar images: the K-Means algorithm and the IsoData algorithm [3]. The main drawback to the application of clustering algorithms is the fact that they tend to distort the image features by imposing a certain shape on the resultant clusters. To alleviate this difficulty relaxation labeling was employed on the clusters to produce new clusters that are less uniform in shape and better reflect the features contained in the marine radar image. The clustering methods showed fairly good results for the segmentation of ice floe features in the marine radar images, a result which is necessary to achieve the goal of automatic tracking. The results of clustering algorithms were not, however, suitable for visual enhancement. More work is required in order to determine the most appropriate feature vector to represent the marine radar images.

With the region splitting algorithm, the marine radar image is

recursively subdivided based on the density of dot-clusters. This results in a good visual enhancement of the image. In order to segment the image and identify ice floe features, the subdivided image must be further processed. Region merging (described below) can be used in order to group together and properly define the regions in the subdivided image.

Region merging [15,16] is a method that groups together pixels or regions that have similar characteristics. This method was developed since it was observed that region splitting yielded good visual enhancement. After applying region splitting to the image, region merging is then used to group together and classify the pixels as distinct regions and entities. In itself, region merging does not produce a very good visual enhancement, but it can be applied to the automatic tracking problem.

The region-based techniques have shown good visual enhancement of the marine radar images. The two methods that were found most suitable for visual enhancement are region splitting and region merging. These same methods have also shown good results in defining regions in the marine radar images for the purpose of automatic tracking.

IV. Conclusions

There were three major categories of algorithms studied, namely, pre/post-processing, edge-based, and region-based algorithms. Many algorithms have been applied for this research—some of them proved to be useful at some stage of the processing, while others did not present any promise at all. The following pre-processing techniques have shown promising results: spatial low pass filtering (median filtering), edge preserving filtering and mathematical morphology. Edge-based segmentation techniques in general did not show promising results for the purpose of automatic tracking but may be useful for the purpose of visual enhancement. The Region-based segmentation techniques, on the other hand, were applied successfully in order to segment the ice floe features in the marine radar images. The region-based techniques that have shown best results are variations of the region splitting and region merging algorithms.

A global result for the visual enhancement of the marine radar image has not been obtained as yet. This will be the goal of future research where many trials of different combinations and variations of the algorithms explained above will be conducted over different radar images. The most appropriate sequence of steps appears to be a noise treatment followed by a combination of region split and merge techniques and finally the application of edge based techniques to highlight the boundaries of the ice floes. An application of the edge preserving and median filters seems to comprise the most effective noise treatment.

Attaining the goal of automatic tracking will also require further research. To date, a combination of region split and merge techniques seems to be the most promising segmentation technique developed for this purpose.

References

- [1] Alparslan, E. et al. "Image Enhancement by Local Histogram Stretching" *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-11, no. 5, May 1981, pp. 376-385.
- [2] Andrews, H. C. et al. "Image Processing by Digital Computer", *IEEE Spectrum*, Vol. 9, no. 7, July 1972, pp. 20-22.
- [3] Duda, R. O., Hart, P. E. "Pattern Classification and Scene Analysis", John Wiley and Sons, 1973.
- [4] Hall, E. L. et al. "A Survey of Preprocessing and Feature Extraction Techniques for Radiographic Image", *IEEE Transaction on Computers*, Vol C-20, Sept. 1971, pp. 1032-1044.
- [5] Haralick, R. M., et al. "On Some Quickly Computable Features of Texture", *Proceedings of the Symposium on Computer Image Processing and Recognition*, Vol 2, EE University of Missouri, Aug 1972, pp. 12-2-1 12-2-10.
- [6] Horowitz, S.L., et al. "Picture Segmentation by a Directed Split-and-Merge Procedure", *Proceedings of the 2nd International Joint Conference on Pattern Recognition*, Copenhagen, Aug 13-15, 1974, pp. 424-533.
- [7] Hummel, R. A. "Histogram Modification Techniques", *Computer Graphics and Image Processing*, Vol 4, April 1974, pp 209-224.
- [8] Hummel, R. A., et al. "On the Foundations of Relaxation Labeling Processes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, no. 3, May 1983, pp. 267-287.
- [9] Justusson, B. "Noise Reduction by Median Filtering", *4th International Conference on Pattern Recognition*, Kyoto, Japan, Nov. 7-10, 1978, pp. 502-504.
- [10] Kohler, R. "A segmentation System Based on Thresholding", *Computer Graphics and Image Processing*, Vol 15, no. 4, April 1981, pp 319-338.
- [11] Mackworth, A. K. "Consistency in the Networks of Relations", *Artificial Intelligence*, Vol. 8, no. 1, 1977, pp. 99-118.
- [12] Nagao, M., et al. "Edge Preserving and Smoothing", *Computer Graphics and Image Processing*, Vol. 9, 1979, pp. 394-407.
- [13] Oppenheim, A.V. et al. "Nonlinear Filtering of Multiplied and Convolved Signals", *Proceedings of the IEEE*, Vol. 56, no. 8, August 1968, pp. 1264-1291.
- [14] Otsu, N "A Threshold Selection Method from Gray-Level Histograms", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-9, No. 1, January 1979, pp. 62-66.
- [15] Pavlidis, T. "Segmentation of Pictures and Maps Through Functional Approximation", *Computer Graphics and Image Processing*, Vol. 1, 1972, pp. 360-372.
- [16] Pavlidis, T. "The Use of Algorithms of Piecewise Approximations for Picture Processing Applications", *ACM Transactions on Mathematical Software*, Vol. 2, no. 4, Dec. 1976, pp. 305-321.
- [17] Prewitt, J. M. S. "Object Enhancement and Extraction", in Lipkin B. S., and Rosenfeld, A., *Picture Processing and Psychopictorics*, Academic, New York, 1970, pp. 75-149.
- [18] Roberts, L. G., "Machine Perception of Three-Dimensional Solids", in Tippet, J. T., et al., *Optical and Electro-Optical Information Processing*, MIT Press, Camb. Mass., 1965, pp 159-197.
- [19] Robinson, G. S. "Edge Detection by Compact Gradient Masks", *Computer Graphics and Image Processing*, Vol 6, no. 5, Oct. 1977, pp. 492-501.
- [20] Rosenfeld, A. "Iterative Methods in Image Analysis", *Pattern Recognition*, Vol. 10, no. 3, 1978, pp. 181-187.
- [21] Rosenfeld, A., et al. "Scene labeling by Relaxation Operators", *IEEE Transactions on Syst. Man and Cyber.*, Vol. SMC-6, no. 6, June 1976, pp. 420-533.
- [22] Serra, J. P. "Image Analysis by Mathematical Morphology", Academic, London, 1982.
- [23] Weszka, J. S. "A Survey of Threshold Selection Techniques", *Computer Graphics and Image Processing*, Vol. 7, 1978, pp. 259-265.

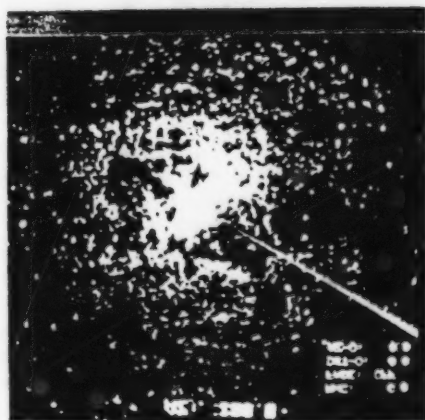


Figure 1 A Typical Marine Radar Image Taken Aboard the Ship.



Figure 2 The Marine Radar Image After Noise Treatment and the Application of Region Splitting.

Logic-Cell Array Implementation of a (31,k) Reed-Solomon Codec and its Applications

by

O.Dravnicks, W.D.Little, V.K.Bhargava
Department of Electrical and Computer Engineering
University of Victoria

Abstract An LCA implementation of an on-line programmable (31,k) Reed-Solomon codec is discussed. Berlekamp's decoding algorithm was modified to allow correction of between one and fifteen 5-bit symbol errors in a thirty-one symbol block. The modified algorithm was tested using the C programming language. The codec was designed using the Xact Design Editor and PC-Silos simulator from Xilinx Inc. running on an IBM PC-AT compatible computer with 5M extended memory. The codec is implemented with Xilinx 3000 series Programmable Gate Arrays. We also describe some applications of this CODEC.

1. Introduction

Reed-Solomon codes are a class of cyclic random error-correcting non-binary block codes discovered by Reed and Solomon in 1960 [1]. The symbols of a Reed-Solomon code are Galois field $GF(2^m)$ elements, where m is a positive integer. An RS code with symbols from $GF(2^m)$ has a block length of 2^m-1 symbols, and can be written to correct t errors, where

$1 \leq t \leq 2^{m-1}-1$. A t -error correcting RS code of block length $n = 2^m - 1$ has $2t$ parity check and $n - 2t$ information symbols per block, and a minimum distance $d = 2t+1$. Decoding an RS code involves four basic steps[2][3]:

- (1) computing the syndrome components S_i , $i = 1, 2, \dots, 2t$
- (2) determining the error-location polynomial $\partial(X)$
- (3) finding the roots of the error-location polynomial $\partial(X)$
- (4) finding and correcting the error at each error location

(1) The symbols r_k in the received block of an RS code are the coefficients of the received vector

$$r(X) = r_0X^0 + r_1X^1 + r_2X^2 + \dots + r_{n-1}X^{n-1}$$

where k is the position of the symbol r_k in the received block. The received vector $r(X)$ is the sum:

$$r(X) = t(X) + e(X)$$

where $t(X)$ is the transmitted vector and $e(X)$ is the vector of errors introduced during transmission. The $2t$ syndrome components are found by substituting α^i into the received vector $r(X)$ for $i = 1, 2, \dots, 2t$:

$$S_i = r(\alpha^i) = t(\alpha^i) + e(\alpha^i) = e(\alpha^i) \quad (1.1)$$

where α is a primitive element of $GF(2^m)$.

(2) The decoder must find the locations of the errors introduced by transmission; that is, the non-zero coefficients of $e(X)$. Directly solving the system of equations (1.1) is difficult; an alternative approach is to introduce a polynomial, the error-location polynomial $\partial(X)$, whose roots are the locations of the transmission errors.

$$\begin{aligned} \partial(X) &= (1 + \beta_1 X)(1 + \beta_2 X) \dots (1 + \beta_\mu X) \\ &= \partial_0 + \partial_1 X + \partial_2 X^2 + \dots + \partial_\mu X^\mu \end{aligned}$$

where β_i is the location of the i th error and μ is the number of errors introduced during transmission. A number of methods exist for determining the error location-polynomial, $\partial(X)$ [3]; Berlekamp's iterative method [4] was used in the programmable (31,k) Reed-Solomon decoder.

(3) The roots β_i of the error-location polynomial must be found. This can be done either by the Chien [3] search or by substituting all the elements α^j in $GF(2^m)$ into the error locator polynomial and noting the elements which give $\partial(\alpha^j) = 0$.

(4) Finding the errors at the error locations requires solving the equations:

$$S_k = Y_1 \beta_1 + Y_2 \beta_2 + \dots + Y_\mu \beta_\mu \quad \text{where } k = 1, 2, \dots, 2t$$

for the μ errors $Y_1, \dots, Y_\mu$. An easy way of doing this is to first find the function

$$\begin{aligned} Z(X) &= 1 + (S_1 + \partial_1)X + (S_2 + \partial_1 S_1 + \partial_2)X^2 + \dots \\ &\quad + (S_\mu + \partial_1 S_{\mu-1} + \partial_2 S_{\mu-2} + \dots + \partial_\mu)X^\mu \end{aligned}$$

The μ errors Y_i can then be found using

$$Y_i = \frac{Z(\beta_i^{-1})}{\prod_{j=1, j \neq i}^{\mu} (1 + \beta_j \beta_i^{-1})}$$

The transmitted vector $t(X)$ can then be found by adding the error vector $e(X)$ to the received vector $r(X)$

$$t(X) = e(X) + r(X).$$

2. The (31,k) RS Decoder Algorithm

Based on the above formulation, an algorithm was written to code and decode a t -error correcting RS code, where t is user programmable and $1 \leq t \leq 2^{m-1}-1$. The algorithm consists of steps

(1) to (3) as given in the introduction, plus two steps from step (4) above: determining $Z(X)$ and finding and correcting the transmitted errors. Each step of the algorithm is dependent upon earlier steps for intermediate results, but runs independent of earlier steps once those intermediate results are received. The programmable RS decoding algorithm was initially written and tested using the C programming language. After testing it was written in terms of hardware instructions. All references to the number of algorithm operations and bit rate refer to the hardware implemented algorithm.

3. Implementation Options

Implementation of the algorithm requires decisions to be made on:

- (1) which basis to use to represent the Galois field and the design of the Galois field arithmetic units
- (2) the number of bus lines and registers to provide for logically independent operations
- (3) the amount of pipelining to be used.

These decisions are made on the basis of how they will effect the trade-off between speed and hardware requirements. These decisions are not independent; decisions made on different aspects of the implementation effect each other.

Basis and Galois Field Arithmetic

The Galois field elements can be represented in either vector or power notation. For example, the element α^{27} from $GF(2^5)$ may be represented as the vector:

$$\alpha^{27} = 1\alpha^0 + 1\alpha^1 + 0\alpha^2 + 1\alpha^3 + 0\alpha^4 \\ = (1\ 1\ 0\ 1\ 0)$$

or as the integer

$$27 = (1\ 1\ 0\ 1\ 1).$$

The advantages and disadvantages in a representation lie in the implementations they allow for Galois field arithmetic, and in the complexity of the hardware needed to implement the decoder in the representation. The comparison of vector and power representation implementations of Galois field arithmetic is dependent on the size of the Galois field, and on the medium on which the arithmetic unit is implemented. In this paper only Xilinx[5] implementations of arithmetic on the Galois fields $GF(2^5)$ will be discussed; the number of configurable logic blocks (CLB's) of the LCA needed to implement arithmetic on the Galois field $GF(2^5)$ for vector and power representations is given in Table 3.1. The power representation has the advantage of allowing the Galois field elements to be used as incremental counters; this simplifies the design of some parts of the decoder.

The algorithm requires Galois field addition, multiplication and inversion. However the number of operations to completion of the fifth step in the algorithm is a third-order equation in terms of the number of correctable errors when the only Galois field arithmetic units are addition, multiplication and inversion. The inclusion of a unit which can raise a Galois field element to an integer power

(integer exponentiation) reduces the number of operations to the fifth step to second order. The ratio of bit rates to the fifth step with an exponential multiplier to the fifth step without an exponential multiplier ranges from about 1.1 at low correctable error rates to about 3 at high correctable error rates.

Galois Field Adders: In the vector representation, Galois field addition is simply integer addition modulo 2. On the Xilinx 3000 series LCA[5] a vector representation adder requires one configurable logic block for every two bits of the field size, plus input registers, for a total of 10 CLB's for a $GF(2^5)$ adder. The easiest implementation of power representation addition on $GF(2^5)$ on the Xilinx 3000 series LCA is to translate from power to vector representation, add mod 2, and translate back to the power representation. The five function logic of the Xilinx 3000 series LCA allows all of this to be done in the input and output registers, and requires 18 CLB's.

Galois Field Multipliers: Choosing the normal basis for the vector representation allows the Massey-Omura[6] multiplier to be used for the Galois field multiplication. For the $GF(2^5)$ multiplier, the least-complex parallel Massey-Omura multiplier requires 20 CLB's, plus 8 CLB's for the input and output registers. Galois field multiplication using the power representation is equivalent to integer addition modulo 2^m . For the $GF(2^5)$ multiplier 12 CLB's plus 5 CLB's for the input registers are needed. The saving in space for the power representation Galois field multiplier is at the cost of a second operation required by the addition modulo 2^m . This second operation can be done in parallel with another operation in the case of the programmable RS decoding algorithm, and so does not effect the overall speed of the decoder.

Galois Field Inverter: Inversion over the field $GF(2^5)$ is identical for either vector or power representation using the five function logic of the Xilinx 3000 series PGA. The total space requirement is 8 CLB's.

Galois Field Integer Exponentiation: This can be done by converting from vector to power representation. In power representation this uses 20 CLB's plus 8 CLB's for the input and output registers. The translation between vector and power notation adds 8 CLB's to the vector notation.

Representation	Implementation Size in CLB's			
	Adder	Multiplier	Inverter	Exponentiation
Vector	10	28	8	36
Power	18	17	8	28

Table 3.1 CLB's for Galois Field Arithmetic

Memory and Bus Lines

During the decoding process the algorithm requires storage space to hold results of intermediate calculations. The intermediate results include results used only in a given step and results to be used in steps after the step in which they are calculated. If there is pipelining the received RS code blocks must also be stored from the first to the last stage in the pipeline. The storage space in bits of

results calculated in previous steps required by each step is given in Table 3.2. Also given in Table 3.2 is the number of CLB's needed by each step of the algorithm, including arithmetic units but not including results calculated in previous steps. The Routing limitations may give rise to more CLB's than that given by this table.

Step	CLB's Used	Bits from Previous Steps
1. Syndrome	95	155*
2. $\partial(X)$	310	155
3. Locations	130	85
4. $Z(X)$	120	240
5. Errors	125	320*

* includes 155 of received code block

Table 3.2 CLB's Used and Previous Bit Storage per Step

Dedicating CLB's to each intermediate variable simplifies the design and speeds up the decoder, but at the cost of space. The implementation of the 5-bit buses required by the $GF(2^5)$ can be done either as common buses using Xilinx's tri-state capabilities, or as buses between each set of communicating decoder components or storage block. It was found that the former scheme allowed for tighter routing.

Pipelining

The independence of the five steps in the decoding algorithm allows a trade-off to be made between the decoder speed and the amount of hardware needed for implementation.

Speed: The number of operations each step requires to complete as a function of the number of correctable errors is given in Fig.1. Not only does each step require a different number of operations to complete, but the number of operations varies among steps either linearly, as a square, or as a cube with respect to the number of operations to completion.

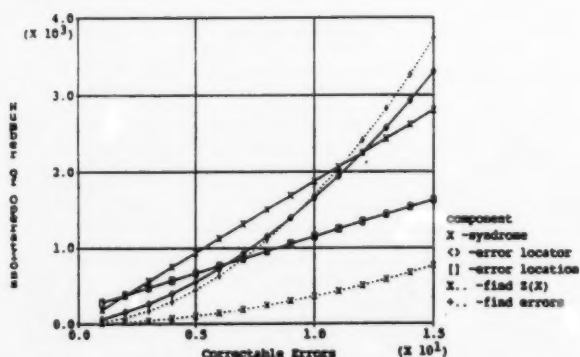


Fig.1 Component Operations vs Correctable Errors

Since the decoder must finish decoding each received block in the same amount of time, the pipelining should be made so that each stage in the pipeline takes approximately the same number of steps. There are six pipelining options which meet this requirement:

- 1) no pipelining
- 2) two stage
 - i) calculate syndrome & find error-location polynomial
 - ii) find error locations & $Z(X)$ & find and correct errors
- 3) three stage
 - i) calculate syndrome & find error-location polynomial
 - ii) find error locations & find $Z(X)$
 - iii) find and correct errors
- 4) four stage, option 1
 - i) find syndrome
 - ii) find error-location polynomial
 - iii) find error locations & find $Z(X)$
 - iv) find and correct errors
- 5) four stage, option 2
 - i) find syndrome
 - ii) find error-location polynomial
 - iii) find error locations
 - iv) find $Z(X)$ & find and correct errors
- 6) five stage (fully pipelined)

The overall number of decoder operations for each pipelining option as a function of correctable errors is given in Fig.2.

Hardware: Each step uses a different number of CLB's to carry out its function, as given in Table 3.3. The received block for each stage in the pipeline must also be saved and passed on from the first to last stage, as well as some intermediate calculations. Hardware is minimized if steps are combined which share intermediate calculations. The coordination of the stages and the shared intermediate calculations also becomes more complex as more stages of pipelining are used.

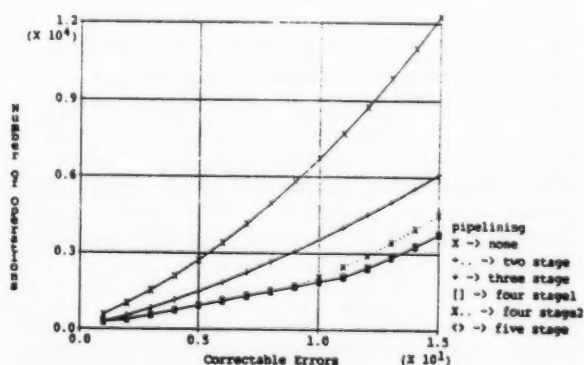


Fig.2 Pipeline Options vs Correctable Errors

Pipelining Option	Storage Required in Bits	
	Intermediate Results	Received Block
None	0	310
Two Stages	240	465
Three Stages	395	620
Four Stages#1	480	775
Four Stages#2	575	775
Five Stages	625	930

Table 3.3 Storage Requirements for Pipelining Options

4. Applications of the CODEC

The (31,k) CODEC can be modified to form a (16,12) CODEC which is a standard for Advanced Train Control systems[7]. This may be accomplished by considering a (31,27) code and then shortening it by 15 symbols to obtain a (16,12) code. The data rate requirement is 4.8 kbps with future upgrades to 9.6 and 19.2kbps.

With $k=15$, we have a (31,15) CODEC which is a standard for Joint Tactical Information Distribution System (JTIDS)[8]. The data rate requirement for this system is 57.6 kbps.

The CODEC is ideally suitable for Meteor Burst Communication Systems[9]. Such systems have recently been proposed for data communications from trucks to dispatch centres.

5. Conclusions

It was decided to implement the (31,k) RS decoder algorithm on the Xilinx 3000 series LCA's with five stage pipelining. power representation of the Galois field, and storage of intermediate results separate from the stages themselves. The five-stage pipelined implementation operates at the same bit rate as the four-stage-option1 pipelined implementation, and requires 205 bits of extra storage. It has the advantage, however, of being totally modular. The handshaking signal between the decoder stages and the intermediate storage component is designed to be general, so that the intermediate storage component may be implemented as RAM or in Xilinx. Each of the five stages is implemented on its own Xilinx chip. If modularity is not considered important, more than one stage could be put on the larger chips of the Xilinx 3000 series LCA's[5]. The use of larger chips increases the overall cost, as the cost per CLB ratio increases as larger chips are selected. Preliminary results suggest a maximum Xilinx chip clock rate of 5 MHz for the complete range of correctable-errors; the resulting bit rates as a function of the correctable errors are given in Fig 3.

The present CODEC does not correct erasures and future work may be directed in this area.

References

- [1] I.S. Reed and G. Solomon, "Polynomial Codes over Certain Finite Fields", *J.Soc.Ind.Appl.Math.* 8,pp.300-304, June 1960.
- [2] S. Lin and D.J. Costello Jr., *Error Control Coding*, Prentice-Hall, N.J. 1983.
- [3] A.M. Michelson and A.H. Levesque, *Error-Control Techniques for Digital Communications*, John Wiley & Sons, 1985.
- [4] E.R. Berlekamp, *Algebraic Coding Theory*, McGraw-Hill, New-York, 1968.
- [5] *The Programmable Gate Array Data Book*, Xilinx, Inc., San Jose, C.A., 1988.
- [6] C.C. Wang, T.K. Truong, H.M. Shao, L.J. Deutsch, J.K. Omura and I.S. Reed, "VLSI Architectures for Computing Multiplications and Inverses in $GF(2^m)$ ", *IEEE Trans. Comput.*, Vol. C-34, pp 709-717, Aug. 1985.
- [7] V.K. Bhargava, T. Le-Ngoc, Y.R. Shayan and A.U.H. Sheikh, "Control Coding Systems for Mobile Communications with Applications to Advanced Train Control Systems (ATCS) and other UHF/VHF Mobile Systems", Transport Canada Publication No. 9596E. Prepared by Binary Communications Inc. under DSS Contract No. 10SB.T8200-7-7516, January 1989.
- [8] R.E. Ziener and R.L. Peterson, *Digital Communications and Spread Spectrum Systems*, New York : MacMillan, 1985.
- [9] V.K. Bhargava, "Communications and Transportation", technical paper prepared for Transport Canada under their "Visiting Expert" program, July 1989.

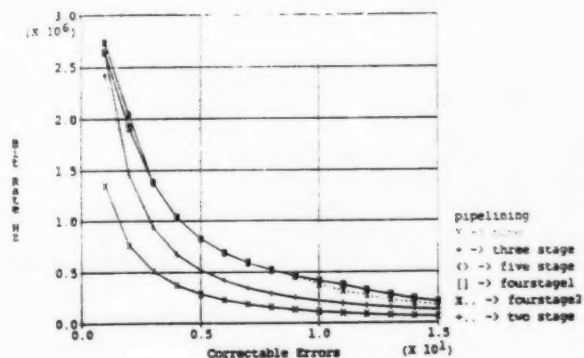


Figure 3 Bit Rate of Pipeline Options vs Correctable Errors

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

AUTOMATED BAGGAGE THREAT DETECTION

Thomas E. Feather
Adrian Lee-Kwen
Ling Guan
Stuart J. Berkowitz

Maurice Audette

Array Systems Computing, Inc.
5000 Dufferin Street
Suite 200
Downsview, Ontario
M3H 5T5

Transportation Development Centre
Complexe Guy Favreau
200 Dorchester Blvd. West
Suite 601, West Tower
Montreal, Quebec
H2Z 1X4

Abstract

A prototype system to detect potential threats as seen in x-ray images of passenger carry-on baggage is described. The initial version concentrates on identifying handguns in real time. Objects found in the x-ray image are classified into three categories representing varying degrees of potential danger. At the highest (red) level, the system attempts to detect all flat, non-occluded handguns. At the second (yellow) level, it attempts to detect all other non-occluded guns and areas that could hide a flat gun. All remaining potentially dangerous objects, for example that could conceal a badly oriented gun, are detected at the lowest (blue) level. Classified objects are highlighted in the x-ray image with the colour that corresponds to the object's threat level.

The system is based on low-cost computer hardware consisting of an 80386-based PC, a display board, an interface board to the x-ray machine, and a board with multiple high-speed processors. A colour graphics display with a touch-screen is employed for the operator interface. A large selection of image manipulation functions are available to the operator via the touch screen.

The system can also detect other dense threats such as pipe bombs, grenades, and bottles filled with liquid. It is currently being extended to detect additional objects such as knives, scissors, electronic devices, and batteries.

Introduction

Transport Canada (TC) owns and operates some 90 airport x-ray screening units which are used to screen passenger carry-on baggage. An operator passes bags through the machine and monitors a scrolling gray-scale display of transmitted x-ray intensities. He attempts to identify bags that may contain restricted items (or "threats"), and subjects such bags to hand searches.

A 140 KVcp x-ray tube illuminates each object as it passes along the conveyer belt. A linear array of 480 diodes sample the transmitted radiation with 8 bits of resolution. Every 8.3 milliseconds, the 480 diode readings are written to the x-ray machine memory and scrolled onto the operator's screen. A typical bag takes a total of 2 - 4 seconds to image.

The resulting x-ray images provide good

definition of the shape of dense objects, such as a handgun, in the absence of other dense occluding objects. When occlusion by objects of similar or greater density is present, it can be very difficult to identify the individual objects. Although human operators are very good at recognizing shapes in the x-ray images, it requires a high degree of attentiveness and concentration. As a result, operators can miss threats that are readily recognizable.

This led to the concept of a Computer-Assisted X-Ray Screening System (CAXSS) to assist operators in screening bags. CAXSS is intended to identify and bring to the operator's attention potentially dangerous objects and regions that could hide dangerous objects. For an initial prototype system, it was decided to concentrate on handguns. The goals for the prototype were detection of handguns with 90% probability and a false alarm rate of less than 10% with a throughput rate of at least 12 bags per minute using relatively low-cost hardware.

There are many varieties of handguns with widely varying shapes. In addition, a gun can be placed in any orientation inside a bag, and can be occluded by metal objects. A gun with the handle aligned vertically is very difficult to recognize. Furthermore, gun handles are much lighter than the gun frame and barrel, and are easily occluded and generally difficult to segment out of an image. Revolver handles are particularly difficult since the metallic parts of the handle are very thin. Moreover, it must be assumed that anyone attempting to pass a gun through an x-ray machine will attempt, as far as possible, to disguise that gun. It can, therefore, be assumed that any approach based purely on looking for a typical flat gun shape would be essentially useless in practice.

There are many problems with the current generation of x-ray machines that also must be contended with. For example, the x-ray image of a completely uniform piece of metal shows very large and near-random variation both across the diodes and within each diode. Although there is a tendency for objects to appear less dense at the edge of the image, this is not consistent. The machine also shows significant variation over time. Long, thin objects (e.g. a long gun barrel) also tend to appear less dense when aligned across the conveyer belt than when aligned along it. Overall, these problems mean that potentially valuable density information is very difficult to interpret reliably.

System Description

System Architecture

The CAXSS is built around an 80386-based PC. A 386 system was chosen to minimize hardware costs after determining that it could support the required data rates. Most of the computation is performed by an embedded board containing several high-speed general purpose processors. This approach was chosen in order to allow maximum flexibility and expandability, while keeping costs to a minimum. Alternatives, such as floating-point array processors and bit-serial parallel processors, were considered and deemed unsuitable, primarily due to excessive cost.

Data are obtained directly from the x-ray system digital memory via a custom interface. Data are pre-processed by the PC and passed to the high-speed processors, which process the data and return the results to the PC to be displayed. Data collection, pre-processing, transfer of data to and from the high-speed processors, and processing of the data are all performed in parallel. Processing of the data is performed on 2 to 4 high speed parallel processors. Typically, results of processing the first part of a bag can be completed and displayed before data collection for the bag is completed.

Object Segmentation

One of the most computationally intensive activities is segmenting the image into distinct objects. After analyzing many images in detail, it was decided to use a two-stage technique. The first stage employs simple thresholding and region growing. This is performed in a very efficient manner on bit-packed data to minimize the execution time. When a potentially interesting object is segmented, it is passed to a high-speed processor to perform a second more detailed segmentation. The second stage employs two-level thresholding. Pixels below the low threshold are termed "core" pixels and are included in an object during region growing regardless of other considerations. Pixels with a high gradient value are also included in order to ensure a complete segmentation of gun handles and triggers.

Feature Extraction

A large number of features are computed for each segmented object. Basic features include area, density, mass, variance, diameter, and perimeter. These are computed both for the entire object and for the "core" of the object consisting of the densest pixels. The core values are particularly useful for occluded objects, since they are relatively insensitive to occlusion, except by extremely dense objects. Various simple derived ratios are also employed such as diameter/area and compactness.

For recognition of gun shapes, a perimeter chain-coding technique is employed (reference [1]). After smoothing, this technique is used to identify the longest near-straight line in the perimeter, which is hypothesized to be one side of a possible barrel. The algorithm tolerates deviations from a straight line so as to find the edge of almost all gun barrels in typical bags. This

is then used to determine a maximal rectangular barrel. Having found a possible barrel, a potential handle is found in a similar manner. For the case of non-occluded guns, this technique finds the barrel and handle of a gun quite well in almost all cases where we are able to completely segment the gun. The barrel of a gun with the handle not visible can typically also be identified correctly. After finding a potential barrel-handle, we compute various derived statistics such as the barrel length, width, mass, density, percent coverage of the object by the barrel and handle, and percent overlap of the barrel and handle. Experimentation with large samples of non-gun objects shows that few non-gun objects exhibit a barrel-handle with properties similar to those of typical barrel-handles found for real guns.

Many other features were investigated and rejected. The time constraint greatly restricts what techniques are feasible and also restricts the number of such techniques that may be employed. The initial attempt at shape recognition employed a medial-axis transform technique to derive the skeleton of the object and then examine various properties of that skeleton, such as whether or not it exhibited an "L" shape which is characteristic of a handgun. Although this technique allowed identification of guns quite well, it also produced an unacceptably high false alarm rate.

Object Classification

Objects are classified into three threat levels of increasing severity. The three levels are designated blue, yellow, and red, with blue being the lowest and red the highest. The goal is to colour red any non-occluded flat guns where the handle is clearly visible. We then attempt to colour all other non-occluded guns yellow as well as all flat occluded guns. Finally, the system attempts to colour blue all other objects that could potentially hide a gun. All occluded guns, no matter how bad the occlusion, are intended to be identified as at least a blue level threat.

Our classification technique employs a combination of discriminant functions and heuristic functions of the extracted features. Objects are classified blue based on passing a first level linear discriminant function or based on other heuristic criteria. The first level function employs primarily occlusion-insensitive features such as "core" measurements. Objects of a sufficient size and mass are also classified blue even if they are not caught by the discriminant function. The functions are defined so as to guarantee classifying virtually all guns as blue, without considering the associated false alarm rate.

An object is coloured yellow only if it was coloured blue and it passes a second level discriminant function. This function employs the barrel-handle feature, which is only computed for objects that are at least blue. A three-class function is used to discriminate non-gun objects vs. flat guns vs. other non-occluded guns. The function at this level is computed so as to produce an acceptable false alarm rate at the yellow and red levels.

Finally, an object is coloured red on the basis of two heuristic functions that employ the barrel-handle feature. The functions estimate how closely the object resembles a flat gun based on the barrel-handle feature and various derived features. If a blue level threat resembles a gun very closely, then it is coloured red. If a yellow level threat resembles a gun, possibly with somewhat less certainty, then it is also coloured red. The function parameters are tuned to provide a very low false alarm rate.

Operator Interface

As data are received from the x-ray system, they are displayed on the operator's screen exactly as received. As soon as an object has been segmented and classified, that object is highlighted via a colour-coded outline of the greyscale image. If an object is classified as either yellow or red, an audible alarm sounds to alert the operator. When a bag has been completed, a window is drawn around the highest-level threat (if any) in the image in order to allow the operator to examine the image further. Optionally, a default sequence of actions (e.g. zoom and histogram equalize) can be performed to the highest level threat automatically.

The CAXSS operator interface employs a touch screen in order to avoid the use of a keyboard. The operator has a set of supervisor-configurable buttons to perform operations such as cycling through objects, zooming, false colouring, histogram equalization, inversion, median filtering and various other operations. Four macro buttons can also be defined by a supervisor to perform any desired combination of functions.

CAXSS has three basic modes of operation: operator, supervisor, and analyst. Passwords are required to enter supervisor and analyst mode. Operator functions are completely defined by a supervisor and may be severely restricted. A supervisor has unrestricted access to most image processing functions as well as the ability to configure the operator mode, calibrate the system, and modify a restricted set of parameters. All supervisor functions are accessible through the touch-screen interface, so that no keyboard would ever be required in operational use. Analyst mode provides access to all system parameters as well as functions used in algorithm development. For example, histograms and cross-sections can be plotted, and the exact barrel-handle found by the system can be displayed on the screen.

Results

Testing Procedures

The results of the work can only be assessed by examining the statistical performance of the system on real data. A large sample of images containing handguns inside and outside bags was collected to assess the detection performance. A total of 16 handguns including 6 revolvers and 10 auto-loaders was used. The sizes and shapes of the 16 guns varied over a wide range. Images were taken of the guns on two different x-ray machines.

The gun images were divided into three classes: flat non-occluded, non-flat non-occluded, and occluded. About 200 flat gun images were used as well as about 100 of each of the other two classes. A flat gun is defined to be a gun with the handle clearly visible that does not overlap or touch any other dense material in the bag. The non-flat guns include primarily those with the handle pointed down in the bag so that only the barrel is visible. The occluded guns included many highly occluded guns as well as only marginally occluded guns. This latter sample is the least well-defined and the results can only be considered representative of the particular sample used.

The false alarm rate performance was assessed using several hundred images of real bags collected randomly at the Ottawa airport using a third x-ray machine.

Test Results

The results are shown below. The percentage of guns and bags containing no guns ("Non-Gun Bags") falling into each threat level is shown. In addition, an efficiency measure is shown. We define efficiency to be a function of the Probability of Detection (PD) and the False Alarm Rate (FAR) as shown below.

$$\text{Efficiency (E)} = \frac{\log(\text{FAR})}{\log(\text{PD})}$$

At a given level, the PD is the detection rate at and above that level, and the FAR is the alarm rate on airport bags at and above that level. For our goals of 90% detection with 10% false alarms, the efficiency is about 32. It appears that, within limits, the PD and FAR for a given level/class can be varied without changing the efficiency significantly.

	Colour			
	None	B	Y	R
Flat Guns (N=200)	0%	2%	3%	95%
		E=99	E>99	E>99
Non-Flat Guns (N=107)	2%	10%	81%	7%
		E=76	E=18	E=1
Occluded Guns	1%	9%	74%	16%
		E=99	E=23	E=3
Non-Gun Bags (N=357)	76%	14%	9%	1%

Figure 1: Detection Rates

The system is able to segment and recognize flat guns with extremely high efficiency, finding 95% at the red level with less than a 1% false alarm rate. No systematic problems were observed. The barrel-handle algorithm correctly identified the barrels of virtually all flat guns, and correctly identified the handles about 95% of the time. Although it is relatively easy to concoct gun-like shapes that cause false alarms in the lab, we found that only 2 of 357 bags passing through an airport x-ray machine caused a red false alarm. The only flat guns that were not coloured red or yellow were at least caught

at the blue level. The problem in each case was an incomplete segmentation of the handle of the gun. Each of the two guns that were coloured blue in some images were coloured yellow or red in other images, so the system did not reliably miss any of the test guns.

Detection of non-flat guns was also very good, approaching 90% detection at the yellow and red levels with less than 10% false alarms. It was found that 9 out of the 13 guns classified as either blue or no colour were small guns with the barrel aligned vertically. Detection of guns where the barrel was horizontal was about 96% at the yellow and red levels. This is partly explained by the fact that few guns with a vertically aligned barrel were employed when computing the discriminant functions, although it is not clear that it is possible to recognize these reliably as guns. It is also the case that these are the most difficult for a human to recognize since they have no distinguishing characteristics. In several cases, apparently due to x-ray system variability, they do not even appear to be very dense.

Detection of occluded guns appeared to approach the ability of a human operator to recognize hidden guns at the yellow level. In the particular sample we used, 91% were detected at the yellow and red levels. However, the segmentation was found to be poor in some instances of occlusion. In particular, when a flat gun is touching a dense object, the system considers the gun and the other object to be one object, and typically colours the object yellow or occasionally blue.

It should be noted that, although we treat the alarms on airport bags as false alarms, this is not entirely correct, except at the red level. The system is intended to detect dark regions that could conceal a gun or other threat. In fact, some of our heuristics specifically cause yellow alarms based purely on size and density even if the object doesn't fit any of our gun models. Inspection of the yellow false alarms showed that, in our opinion, virtually all of these could conceivably be non-flat guns or at least conceal a gun. A few appeared to be large bottles filled with liquid, which are considered to be potential threats. In any case, it appeared that they all would require hand searching.

A typical handgun is classified and highlighted within about 2 seconds after it has been imaged. Typical real bags collected from the airport are processed almost as fast as the data are collected, typically requiring less than 1 second after a bag is imaged to complete the analysis. With four high-speed processors, almost all bags are processed in less than the data collection time.

Conclusions and Future Work

The present work demonstrates that effective automated detection of handguns in x-ray images is possible using low-cost hardware. The measured detection and false alarm rates compare very favorably with human operators. Typical detection rates for operators are around 75% for flat guns with 10-15% total alarm rate, although this alarm rate includes alarms related to all threats.

The system's detection of handguns can yet be improved in several ways. The primary improvement would come from employing more advanced segmentation algorithms to ensure that all guns can be segmented properly. This will require more computational power than is currently available in the system. The system can be easily upgraded to make use of additional high-speed processors, although this would obviously increase the system cost. New generations of higher-speed, lower-cost processors also can be easily incorporated when they are available.

Although the processing speed is quite good, an operational version of the system should probably de-couple data collection and display so that the next bag can be collected and processed while the operator examines the current bag image. The speed of the prototype system is such that it could support a continuous stream of bags at the maximum x-ray scanning rate if the data collection time were available for processing.

Work is in progress to add additional threat categories to the system. Relatively dense threats, such as pipe bombs, grenades, and liquid-filled bottles, are already detected by the current algorithms, typically at the yellow level. The perimeter chain coding algorithm has been adapted to identify sharp objects, and preliminary testing on knives and scissors has produced good results. The blades of most relatively large knives are identified correctly, with a reasonable false alarm rate.

It is also hoped that the system can be extended to help detect explosive devices. A texture algorithm is being developed to recognize "complex" areas, typically signifying electronic items, which are good hiding places for bombs. Preliminary testing shows promising results. Algorithms are also being developed to recognize batteries and wires with a reasonable false alarm rate. It is expected that the presence of multiple coincident bomb indicators, such as batteries with wires attached, will be typical of x-ray images of several classes of explosive devices while being sufficiently uncommon in normal baggage. To augment this approach, the use of a dual-energy or backscatter x-ray device is also being considered. These allow an approximate determination of the material composing each object.

References

- [1] R. Safaei-Rad, B. Behabib, K. C. Smith and K. M. Ty, "Position, Rotation, and Scale Invariant Recognition of Two-Dimensional Objects using a Gradient-Coding Scheme", Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Vancouver, B.C., 1989, pp. 306-311.

A GRAPHIC INTEGRATED SOFTWARE PACKAGE FOR ROAD MANAGEMENT

A.Bedetti (1), R.Daverio (2), B.DiStefano (3), E.Distefano (3).

Abstract

This paper describes the GSI (by SOFOS S.r.l.) road management integrated software package. The GSI consists of two parts:

- a computerized street map;
- a data base of static structures correlated to the street map.

GSI has been implemented using AutoCAD (by Autodesk Inc.), AutoLISP (by Autodesk Inc.), DBIII Plus, and a number of custom developed routines coded in the C programming language. The first generation has been developed for use under MS-DOS. A UNIX version is currently under design.

The computerized street map can be:

- imported from an existing geographic data base containing standard mapping information, as long as this can be converted into DXF or DXB file formats;
- digitized from existing maps by use of either CCD cameras or scanners;
- manually digitized by an operator.

The use of existing mapping information from an electronic geographic data base is definitely the most time and cost effective solution. Lacking such information, the use of CCD cameras or of scanners is the next best choice. However, sometime the information available through these two systems may be excessive and require needlessly large amount of RAM and mass storage. Thus, it can be justifiable to use an operator who can manually digitize the required information. Often a combination of the three methods is required.

To allow high speed regeneration of images (display of new images), the map is divided into twenty "sections". Each section is divided into nine "areas". Each area has dimensions proportional to a typical CRT screen (3/4 height, 4/4 width). In this way it is possible to store in different parts of RAM information related to different sections and areas. Thus, it is possible to switch from the display of one area to another, without any disk access and without any recalculation of coordinates. The resulting speed improvement is beyond one order of magnitude.

Within each area the important elements are "points". They are beginning, end, and intersection of streets. Only points and the lines joining them (streets) appear in the data base. A "path" is the set of all points and streets that allow to go from point A to point B. It is possible to define "relevant" points. They are not related in any way to streets. Relevant points identify static elements that are considered important for a specific application (i.e: telephone booths, gas stations, bus stops, police stations, etc.).

By using the computerized street map and the data base it is possible to derive information of dynamic nature (i.e: routes). GSI contains a number of built-in algorithms that allow the automatic identification of the shortest route and an approximate calculation of the time required to cover a given route. As GSI does not have information about street numbers, all routes have to be identified by the fixed points that are already in GSI data base of the specific application.

A convenient use of GSI has been to generate optimized daily route sheets for drivers. This use may be attractive to delivery companies and school bus operators. The availability of suitable small computers may allow the use of GSI on vehicles.

All GSI functions are menu driven and all menus are of the AutoCAD style.

Overview

The acronym GSI stands for "GESTIONE STRADARIO INTEGRATO", that is Italian for "Integrated Road-Map Management System". This name has been chosen because the first release of GSI has been customized for the city of Bologna, in Italy, (141 square kilometers).

The GSI road management integrated software package is a general purpose graphic application for the management of road related public services and of road based resources. GSI requires some customizing for every new application. An experienced engineer (or other person of comparable education) must customize GSI for a specific application and eventually for a specific end user. After this operation is completed, the end user can operate the customized version of GSI.

GSI has been developed using AutoCAD, AutoLISP, DBIII Plus, and a number of custom developed routines coded in the C programming language. The choice of AutoCAD as one of the basic building blocks has determined most characteristics and modes of operation of GSI. For instance, GSI has a very wide flexibility when importing a map.

-
- (1) SOFOS S.r.l.
Via Callegherie, 13
40026 Imola
(Bologna), Italy
Tel. (542) 23600
- (2) Technisoft
Via Po, 19
21100 Varese, Italy
Tel. (332) 22 45 41
- (3) Nuptek Systems Ltd.
15 McMurich St. (suite 1002)
Toronto, Ontario M5R 3M6
Tel. (416) 925 7231
FAX (416) 975 0759

This is so because all third party software packages available for AutoCAD can be used. Even other packages, not related to AutoCAD, as long as they can generate DXF or DXB files, can be used.

As with other software packages [1] [2], the data base containing the computerized street map can be created in at least three ways:

- 1) it can be imported from an existing geographic data base containing maps, already in digital form;
- 2) it can be digitized from existing hard copy of traditional maps by means of either CCD cameras or scanners;
- 3) it can be digitized manually by an operator, by means of proper pointing devices [3] [4].

As it can be seen after a quick analysis of the available technologies [2], use of existing mapping information from an electronic geographic data base is definitely the most time and cost effective solution. In the absence of such information, the use of CCD cameras or of scanners is the next best choice [5]. However, sometime the information available through these two systems may be excessive and require needlessly large amount of RAM and mass storage. Thus, it can be justifiable to use an operator who can manually digitize the required information or who can edit the information generated using one of the other two methods. It is conceivable to eliminate unwanted information by programming filters, able to identify redundant information. This is possible because of the data structure of AutoCAD [6] [7] [8].

The first release of GSI has been compatible with the popular MS/DOS operating system. A UNIX version is currently under design.

When using the MS/DOS software environment, GSI requires the following minimum hardware configuration:

- IBM XT or true compatible;
- 512 kilobytes of RAM;
- colour graphic adapter;
- colour graphic monitor;
- one serial RS232-c port;
- 80 columns IBM compatible printer.

Better human interface can be achieved using:

- computer that can provide higher execution speed (i.e: IBM AT, IBM PS2, COMPAQ 386/25 MHz, etc.);
- more memory, as it may be required by the application;
- higher resolution video interface (i.e: EGA, VGA, PGA, etc.);
- higher resolution monitor, as it may be required by the video interface;
- plotter and/or laser printer.

While not necessarily required, use of a mouse, a digitizing tablet, or a touch-pen is highly recommended to improve the user-friendliness of the package.

All hardware and software options that can be used with AutoCAD can be used also with GSI.

It is possible to use GSI together with a mainframe. In this case the GSI software still runs on the personal computer or on the workstation. The function of the mainframe may be to store the database. This mode of operation may allow several operators to work simultaneously on the same database.

If all operators can only display the maps or derive data from it, no problem exists. If two or more operators are allowed to modify the map or the database, some form of authorization and arbitration must be enforced. The most reasonable way of solving this problem is to give each operator permission to modify only one type of information, one layer in AutoCAD jargon. In this way all operators share the same data for display purposes, but every operator has privileged rights for alterations.

Availability of small computers, of lap top models in particular, allows the use of GSI on vehicles, because no special peripherals are required for the minimum system. This mode of operation can be useful for route optimization of radio dispatched fleets. Another use of GSI on board of vehicles is the data acquisition of relevant information.

Also in this case, as in the case of the use of GSI by multiple operators via a mainframe, validation procedures must be in place and must be enforced to avoid introduction of errors in the database.

Concepts and Definitions

In order to improve speed of execution and to minimize memory requirements, a model has been created, that ignores all redundant data, while retaining all important information required by the application. This model is based on a small number of basic conceptual entities:

- drawings;
- drawing entities;
- sections;
- areas;
- points;
- relevant points;
- paths;
- routes.

A short description of these entities follows.

The "drawing" is the file containing all information describing the graphic aspects of the application. The drawing can be generated using any unit (metric, imperial, etc.) and can have any dimension. The drawing file reference system is a cartesian (x,y) system, with the origin (0,0) in the left bottom corner of the screen. By loading a different drawing file it is possible to look at different cities (or boroughs of the same large city) and to work on the related database. Any drawing file, if fully displayed on a CRT monitor, shows exactly the same information that will appear on the hard copy output, either by means of a

printer or by means of a plotter. Full display of drawing files is WYSIWYG ("what you see is what you get"). The drawing file is a typical AutoCAD drawing file, represented as a ".DWG", when displaying the directory of a computer disk.

The "drawing entities" are:

- lines;
- arcs;
- circles;
- text.

All drawing entities are fully AutoCAD compatible.

Every map is divided into "sections", square in shape. These sections are conceptually equal to the pages of a typical book containing a city map. The first page shows the entire city divided into square blocks. By looking at the first page, the user identifies the area of interest and, then, goes on to look at the detailed representation on a different page. The same thing can be done when using the GSI package.

All applications developed so far have consisted of a maximum of 20 sections, but a higher number is conceivable. Every section is divided into 9 "areas". Every area has dimensions proportional to a CRT screen (3/4 height, 4/4 width). This definition allows high speed regeneration (display of new images) on the screen. In this way it is possible to store in different banks of RAM information related to different sections and areas. Thus, it is possible to switch from the display of one area to another, without any disk access and without any recalculation of coordinates. The resulting speed improvement is beyond one order of magnitude.

The exact amount of improvement is hardware dependent. The faster the clock rate and the RAM access time, the faster the switching time from the display of one area to the display of another area. However, this is a software feature and does not use any special hardware.

Within each area the important elements are "points". They are beginning, end, and intersection of streets. Only points and the lines joining them (streets) appear in the data base.

A "path" is the set of all points and streets that allow to go from point A to point B.

It is possible to define "relevant" points. They are not related in any way to streets. Relevant points identify static elements that are considered important for a specific application (i.e.: telephone booths, gas stations, bus stops, police stations, etc.).

A "route" is a set of paths that allows to carry on an activity that requires moving through the city.

It is possible to use the above concepts to process all information related to a given

activity that must be managed. For instance the route of a given garbage collection truck starts from the garage and ends to the garbage processing plant. Alternatively this route could start from the garage and end to the same garage. This route consists of several paths, from garbage container to garbage container, from building to building. The entire garbage collection service can be described by the set of all garbage collection routes. All services offered and managed by a city administration can be represented by the collection of the individual services.

It is possible also to identify the services with the relevant points. For instance the entire garbage collection service can be identified by the relevant point representing:

- garages for garbage collection trucks;
- garbage containers in the street and in buildings;
- garbage processing plants.

However, this view is static in nature and while it can provide an inventory of resources available for the service under consideration, it does not provide any information on the dynamic aspects of how the service is implemented. The route provides the information of dynamic nature, thus, as a consequence on the real time aspects of the operation.

Because of the characteristic of the applications developed so far, the following relevant points have been defined:

- recycling containers for solid waste;
- recycling containers for glass;
- garbage cans at street corners;
- gas stations;
- garbage truck garages;
- public transit garages;
- police garages;
- street signs;
- parking lots;
- trees;
- bus stops;
- pedestrian crossings;
- one way streets;
- driveways.

It is possible to define other relevant point if the need arises. All relevant point of the same type (i.e.: all gas stations) are contained in the same file. The operator may enable and display, in a selective way, relevant points of some types, but not others, as it may be required by the specific operation being carried on at any given time.

For instance, at some point in time the operator may display the files related to the recycling containers for glass together with the base city map. This may be done to prepare a route for the truck collecting such containers. The same operator can then be requested to report on the location of city garages where snow removal equipment is located. All he has to do is to disable the layer where recycling containers for glass are and to enable the layer where the garages containing snow removal equipment are.

Routes can be printed in graphic form on a normal dot matrix type printer if low resolution hard copy is acceptable. If a higher quality drawing is desired, a plotter can be used. Additionally routes can be printed out as a computer listing containing:

- all streets included in the route, in the order in which they are encountered going from starting point to destination;
- length of each path in the route;
- information about average time required to go through each street.

The computer listing of each route is available in two versions:

- a management version, containing enough information for decision making, should the manager decide to modify the computer generated route;
- a user version, containing only the information required for an efficient use of the route.

When the route is printed or plotted in graphic form, it contains only user related information.

The definition of sections and areas has no relationship with the traditional concept of "scale". Sections and area can be defined for any map or set of maps, no matter what scale factor has been chosen to generate the map(s). For all applications implemented so far it has been found that a scale of 1:20000 offers the best trade-off between detailed information and proper disk and memory usage. This scale does not offer the graphic precision of a more detailed scale, but offers several advantages, mainly in terms of speed of execution. If the transition to a more detailed scale is desired, this can be done without changing the algorithms, because GSI works on the above mentioned "points" and not on the graphic representation of the map.

All points and the lines joining them make up a "network" invisible to the operator, but existing at a different level. All algorithms applicable to the map are related to this network.

Two types of data exist:

- geographical data, consisting of as many layers as required by the application (layers here have the meaning defined by AutoCAD and by most other CAD drafting packages);
- descriptive data (database) containing information about all relevant points of the map.

The two types of data are properly cross-referenced.

Depending on the application, the operator may or may not transfer the information into all layers of geographical data.

References

- [1] R. Daverio, "TOPmap Sistema Grafico Integrato per la Gestione di Reti Tecnologiche - Breve Descrizione delle Funzionalità Utente", Technisoft 1988

- [2] R.Daverio, B.DiStefano, E.Distefano, Method of Automated Mapping/Facility Management Suitable for a Personal Workstation", Proceedings of "Canadian Conference on Electrical and Computer Engineering", September 17-20, 1989;
- [3] B. Bradlee, "Input Devices", Cadence Vol. 2, No. 6 (June 1987), pp. 18-21;
- [4] B. Howard, "Point-and-shoot devices", PC Magazine, Vol. 6, No. 14 (August 1987), pp. 95-96;
- [5] Personal Publishing - The magazine for Desktop Publishers, Vol. 3, No. 4, May 1987, (most of this issue is dedicated to scanners, and related topics)
- [6] AutoCAD User Reference Manual, Appendix C, Autodesk Inc., pp. 293-306,
- [7] F. W. Jones, L. Martin, "The AutoCAD Database Book - Accessing and Managing CAD Drawing Information", Ventana Press 1987, Chapel Hill, North Carolina
- [8] R.C. Bradlee, "Programming AutoCAD" Vol. 1, ADSI, Naperville IL.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

EFFECTIVE INTERFACE TO EXPERT SYSTEMS IN TRANSPORTATION

by

Ann Griffith
Takashi Gomi
Karl Mayer
Use Oppacher
Applied AI Systems, Inc
300 March Road, Suite 602
Kanata, Ontario, Canada K2k 2E2

Lewis Sabounghi
Transportation Development Centre
200 Dorchester Blvd West, Suite 601
Montreal, Quebec, Canada, H2Z 1X4

Abstract

The effectiveness of the user interface is the single most important criteria in acceptance or rejection of an expert system. Factors which should be considered in developing an effective interface are discussed. The interface that has been developed for a project for the Transportation Development Center by Applied AI Systems, Inc is explained and discussed taking into consideration the interface requirements for expert systems in general and those in transportation in particular.

1. Introduction

Following some preliminary research which investigated the feasibility of choosing the domain of dispatching for the development of a knowledge based system, Applied AI Systems, Inc (AAI) was awarded a contract by the Transportation Development Centre (TDC) of Transport Canada in Montreal, Quebec to develop a prototype expert system. A branch office of Canadian Liquid Air (CLA) in Hamilton, Ontario agreed to allow developers to observe and interview their dispatcher who is responsible for dispatching the drivers of bulk carriers to deliver liquid gas to their customers.

Through previous experience in another development project for the Ontario Ministry of Transportation, AAI recognized the importance of the user interface in the acceptance of a computer-based system in the field. It is one of five major issues which must be considered in developing fieldable expert systems and the single most important issue in user acceptance. Therefore, considerable emphasis was placed on the development of the interface throughout the project. This influenced the selection of the tool to be used for development and a multi-modal interface was developed for the user.

2. Objective

The objective of the project was to demonstrate the feasibility of knowledge-based systems in the area of dispatching and to introduce some advanced concepts in user interface. Issues that are important in transportation related

applications are identified and attention is focused on the interface which will be most effective in the domain.

Many transportation applications are described as "hands busy/eyes busy". Therefore, in these situations an interface to the computer is required which will take this into account and will be an asset to an already busy dispatcher rather than an aggravation or annoyance.

3. Interface Issues in Expert Systems

Expert systems place special demands on the design of the user interface. These demands are related to the characteristics of the problems that tend to be solved by expert systems. The problems where expert systems are used to provide a solution are often in complex domains where the problem is ill-defined and ill-structured. In general, the more complex the domain and the service offered by the expert system, the better the interface that is required.

It is recognized that the design of the user interface is very important to the success of the expert system. It is not an area that is easily defined or has clear guidelines but rather some principles must be applied and manipulated depending upon the application and the demands of the expert system. The requirements and specifications necessary to build an effective interface are still evolving. The interface itself should develop and evolve throughout the life of the project and is not something that is pasted on after development is complete. In all, a significant portion of development resources must be spent to complete the interface.

The objectives of the interface design are:

- (1) To reduce the mental or cognitive load on the user by mediating the system's complexity;
- (2) Accommodate individual differences for one or more classes of users;
- (3) Produce a positive effect - the goal is that people enjoy using the system; and,
- (4) Meet the objectives within the given time and cost constraints. [Stelzner and Cypher, 1987, 1988]

The interface should give the user a feeling of control while reassuring him that the system is helpful. The user should be able to focus on the domain and the problem and not get bogged down or distracted by the interaction with the computer. The interface design needs to anticipate the user's requirements and meet them in such a way that he does not have to focus on the representation of the problem but the problem itself. A sense of transparency must be present.

There are different levels of experience and expertise in the use of computers and the interface needs to be designed in such a way that a novice to computers can maneuver through the expert system and make use of its functionality and not be frustrated by the fact that he does not know system commands or a lot about computers. For a novice there may be a lot of fear and anxiety about breaking the system; skepticism about the computer ("a machine cannot do as well as I"); and outright rejection ("this machine may take over my job"). The interface can strongly influence the overall impact of these problems. If he is frustrated at the beginning he may refuse to lay his hands on the keyboard and completely reject the system.

The experience of the user in the domain needs to also be taken into consideration. If the system is to be used to train new people in the position then a lower level of knowledge will be assumed than if the system is used by the expert as an assistant for some of the more routine aspects of the job. If the system is to be used by both novice and experienced, then there should be some choices. The experienced user can then bypass or use a more advanced interface so that he does not become frustrated by tedious questions which are necessary for the less experienced. The choice gives the interface a depth. Help screens should be an option and available for those who require them but not routinely presented to aggravate a sophisticated user.

By mimicking the forms in use in the day-to-day operations, the feeling of strangeness is reduced. The user can then see the expert system as an extension of his operations. If he does not have to learn a new form, and where to look for certain information the probability of acceptance of the system is increased.

In filling in forms or answering questions, there are occasions when typing the answer is the best alternative and there are times when a multiple choice menu or pop-up is the best choice. It is important to recognize where each of these should be used. For example, if a typing error will cause an error and many entries would have to be re-done as a result of that error, then it is preferable to have a choice menu where the chance of error is greatly reduced. For someone who is not familiar with the keyboard, having to do a lot of typing can be very tedious, but for an experienced user, pulling down a menu and finding the proper selection can be more frustrating than typing in the answer.

All of the information that is relevant to a particular subtask should be on one screen but it needs to be presented in such a way that it is not overwhelming and confusing. A graphical or tabular means of presentation is easier to read and extracts information and explains a solution clearer than a strictly textual presentation.

4. Domain

The dispatcher of the Bulk Distribution Department at CLA is responsible for assigning drivers for three shifts to deliver product (liquefied gases) to customers in a wide area surrounding the dispatching office. He is responsible for meeting the customer requirements for delivery of the liquid gases Oxygen, Nitrogen, Carbon Dioxide and Argon.

From a computer system already in place in the dispatcher's office a daily printout lists the customers for each gas which require delivery. The system starts predicting that a customer will require delivery four days before it is actually critical that the delivery take place. This gives the dispatcher some flexibility in arranging his delivery schedule and optimizing the deliveries within a zone.

A driver is assigned to deliver a particular gas to customers within a given zone. He does not deliver outside of the zone or combine customers from different zones. This is done for efficiency and has proven to be the most cost effective means of delivery.

In this dispatching process there are many factors which must be considered and given attention before the assignments can be finalized for the drivers. This includes such things as customer limitations - some customers can only receive product at certain times, and because of the nature of the delivery site only certain tractors and trailers can deliver to certain customers; weight regulations - restrictions placed by the Department of Transport; driver preference - senior drivers are allowed to deliver their choice of product in most cases; and union regulations - drivers must be given an equal opportunity for overtime and for the long-distance hauls which they all like.

5. Interface for the TDC Project

With consideration given to the requirements for an effective user interface and recognizing the special needs of the dispatcher in an environment which is always pressured an interface was developed which demonstrated some advanced concepts.

Since the situation may change suddenly, for example a driver may phone in sick at the last moment and the dispatcher is left with insufficient manpower to meet the delivery requirement, and there are constant demands from the customers, any assistance which alleviates some of the work load should be of benefit to the dispatcher. But the system should not be a burden or irritation to him.

It must be designed in such a way that the assistance is available in the least obtrusive way.

The completed prototype incorporates a multi-faceted user interface to demonstrate the flexibility and capabilities of today's technology. The knowledge base was developed using the expert system development tool, NEXPERT. One of the reasons for the selection of this tool for the development was the capabilities to build a custom runtime interface with many of the features which have been identified as important for acceptance by the user.

The visual interface is primarily menu driven for minimum cognitive load to the user, and multiple windows are presented for the retrieval and display of information. Information presented on a screen is relevant to the task at the time and the screens are designed so that the information required for that task is available and the user does not have to recall what had been entered on a previous screen. Where appropriate, the screen interface has been designed with multiple choice selections where the user can select as many or as few choices as required (Figure 1). In other cases the data can be entered by typing or a selection from a pop-up menu will give the same effect (Figure 2). Where data is entered by typing there is an opportunity to review the inputs and an opportunity to make any required corrections. Tables in which entries can be edited are typically used to provide the user with an overview before finalization. This provides an opportunity for changes and/or corrections. The use of color and consistency of keystrokes are all factors which influence the users reaction to the system and have been carefully controlled in the system.

SELECT CUSTOMERS

Trailer	City	State	Zip
WYATONGLO	f	7050	
FRANKSTLO	f	6900	11 24
BISHLEDGO	f	6825	7
PEELHOSPO	f	6867	

Trailer

JOSDHOSPO	f	6867	
ETOHOSPO	f	2911	11 24
MISSHOSPO	f	2933	

City

TEMOLLIO	f	2625	
THERMUBUNG	f	1725	
EXIDEIO	f	1725	07 17
DOCTUOSO	f	1533	

Total selected: 25000
Total Unselected: 21500

SELECT TRAILER

Trailer	Zip	City	State
Tr414	24000	Sedlo	744
Tr413	10000		
Tr409	10000		
Tr405	10000		

end of list

Use (arrow) to move.
(Space Bar) to select/unselect
(B1)-(B) to assign trailer, (F1) to exit

Figure 1

In addition to the entries made through the keyboard, the system can be completely driven by a voice recognition system which is integrated with the software tool. The Dragon Systems, Inc. VoiceScribe system is used for the voice recognition. A limited vocabulary, specific to the expert system user interface, was designed and the voice recognition was "trained" on it. It provides for voice commands equivalent to keyboard inputs which are also active. The questions appearing on the screen can also be posed with a voice synthesis system. The voice synthesis system selected was by AICOM Corporation and the synthesis is accomplished by sending text strings to the output device created by the voice driver. With the installation of the voice recognition and voice synthesis it frees the dispatcher from the computer terminal so that he can retrieve other documents and check on other information while still continuing with the program. These initial capabilities can be further developed and used all together or in any combination depending on the situation and the requirements.

In all, the developed system demonstrates what can be done with an intelligent system which is equipped with a multi-modal user interface.

IF YOU WANT TO SEE REPORTS press (Enter) If not (F1)

To LOOK at reports press (F2)

(arrow) to choose
(Enter) to confirm

To PRINT reports:

Select report
Press (Enter) and then (F3)

(F1) to continue...

REPORTS

SELECTED TRAILERS

GAS TRAILER

ALL TRAILERS

CUSTOMERS

NO REPORTS

Figure 2

6. Conclusions

In general, users react more to the interface to the system than to the functionality of the system. Therefore, emphasis on the interface is very important. Voice input and output is now a viable interface with the computer and there are situations where it is certainly the ideal interface. In areas where one might not always have the hands free for the keyboard or eyes for the screen the use of voice is a viable alternative.

There are still some issues which need to be resolved and require further study. The quality of the voice generated is still less than ideal and there is more cognitive and psychological study required on the ergonomics of the system. A careful design of a dialogue model will also be necessary with the addition of voice capability. The expert system still requires an integration with a richer knowledge base which incorporates more of the real heuristics used by the dispatcher.

References

[Stelzner and Cypher, 1987]

Stelzner, M. and Cypher, A., "User Interfaces". A Tutorial given at IEEE 'The Third Conference on Artificial Intelligence Applications' at Kisimsee, Florida, February, 1987.

[Stelzner and Cypher, 1988]

Stelzner, M. and Cypher, A., "User Interfaces". A Tutorial given at IEEE 'The Fourth Conference on Artificial Intelligence Applications' at San Diego, California, March 1988.

CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING
CONGRES CANADIEN EN GENIE ELECTRIQUE ET INFORMATIQUE
Montréal, PQ, Canada, September 17-20 September 1989

AN EXPERT SYSTEM IN RAILWAY INFORMATION SYSTEM

Al Pozniak, Albert Luk

CN Rail, 935 La Gauchetierre Ouest,
Montreal, Quebec, Canada H3B 2M9.

Abstract :

The railway is one of the most economical means of mass transportation in today's world. In countries where the highway infrastructure is not well developed the railway is the only transportation tool. The main limitation imposed on railway is that railcars can only be moved on rail tracks. Sorting or switching a train of railcars often requires a lot of physical car movements, temporary track space and precious time. To operate an efficient railway we want to keep the car fleet busy and productive by minimizing car idling time and car movement. Computer aided planning tools are becoming more and more important in rail yard operations. With good operational planning, cars can be sorted with least number of movements as such car idling time can be minimized. This paper describes a working computer aided expert system package known as "PLAN" that plans car switching activities for a rail yard.

"PLAN" was developed by the Chinese Railway between 1986 and 1988. It was based on the modified version of YIS (Yard Inventory System) acquired from the Canadian National Railway [1]. Now it is being in use in the busiest yard in China, the Zhengzhou yard. A busy yard like the Zhengzhou yard handles over 20,000 railcars a day. It is the major railcar switch yard in China.

In order to handle the high volume of traffic the car switching movements must be minimized to optimize the car-cycle time. Computer aided expert system in railway yard operation becomes the logical tool to assist the operation. Especially when experienced yard masters are difficult to find and new staff requires long training period to acquire the necessary expertise and knowledge. "PLAN" allows users to ask a defined set of "what if's" kind of questions, and shows users the predicted results up to 16 car switch plans ahead. A yard master can now view the predicted results of a series of car switching activities and then make his best move. The productivity of the Zhengzhou yard equipped with the new system had improved by 15% in the first year of operation.

The CN's YIS hardware, system software and application software were modified for both Chinese input and output. The "PLAN" package was developed upon it to replace the traditional manual ad-hoc queries to gather and digest information for decision making. The system uses a disk file to represent a physical rail track. If a track can hold 100 cars the corresponding file is 100 car records long. There is also a virtual disk file for each track. When the yard has x number of rail tracks there is x number of track files and x number of virtual track files in the system. A stack of 16 history files is maintained to manage a maximum sequence of 16 car switching plans.

The main function of a switch yard can be illustrated in the following simplified diagram.

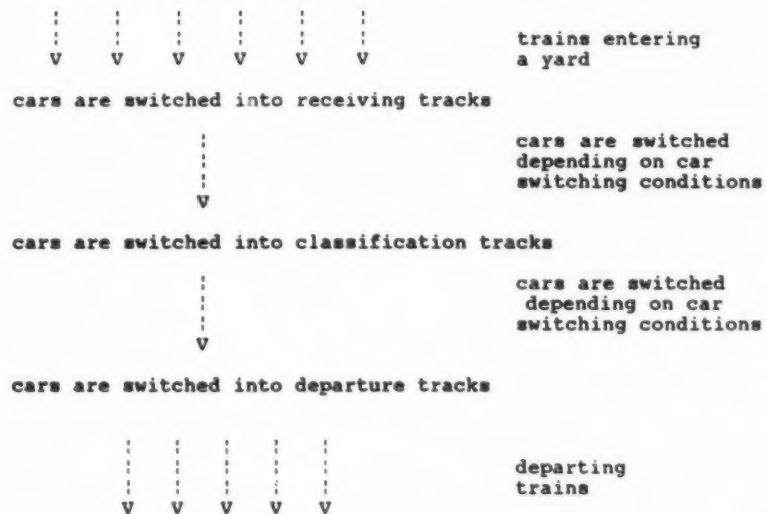


Figure 1. A Simplified Traffic Flow Diagram In A Yard.

As we can see from Figure 1., if the car switching process is slow compared to the incoming traffic volume, and if the receiving tracks do not have enough buffering space, traffic congestion will definitely arise.

Car switching conditions are governed by two major rules - fixed rules and heuristic rules. Fixed rules are those imposed on by the government and engineering for safety reasons. Heuristic rules are the operation rules that minimize car switching time, that build up trains requiring less number of car switches in the next yard. In railway terms, this is referred to as "blocking". There are hundreds or even thousands of these rules, blocking plans, switching plans, etc that a yard master has to remember and relate them to the car switching activities. Obviously a lot of these tedious human processes are error prone resulting in longer car switching time.

The "PLAN" package contains tables of fixed rules defined by transportation users. An example is that if a car contains explosive it should not be placed next to a locomotive or a caboose. Another example is that a tank car carrying petroleum cannot be placed next to a car containing sulphur. Each car record contains codes that describe its contents and characteristics. Also, in the computer system there are train make up catalogs that define criteria to build up a departing train. Criteria are like the minimum and maximum number of cars, minimum and maximum train tonnage, number of cars going in the same direction and etc.

As trains arrive into the yard they are waiting in the receiving yard to be switched and grouped to form departing trains. The yard masters have to prepare a flow through activity plan in advance to avoid bottlenecks during operation. The following scenario best

describes how "PLAN" provides expert assistance to yard masters in yard operation.

The yard master enters a receiving track name R001 (receiving track 1) into the "PLAN" menu on his CRT. PLAN looks at the car records in the virtual file R001'. PLAN checks each car record's content and the last few car records of each virtual classification track to see if it fits the fixed rules requirements. For example, if the car on R001' carries explosives, PLAN will not allow it to be placed on a classification track on which the last several cars also carry explosives or other dangerous commodities. PLAN then checks heuristic rules to place cars going in the same direction on the same classification track. After both fixed and heuristic rules are applied on each car record, PLAN moves car records from R001' to various virtual classification tracks files. Then PLAN generates a track switch plan to the user and displays a car distribution map on a monitor as if the switch plan were actually executed. The yard master examines the computer generated switch plan and the forecasted car distribution map on the monitor. If he agrees, PLAN will store the switch plan aside in a history log file. Otherwise he can modify the switch plan based on his reasonings and knowledge.

If the switch plan were modified, PLAN will first undo and pre-empt the switch plan's result by moving back car records from various virtual classification tracks files back into the original virtual R001' file. Then PLAN has to redo the result of the modified switch plan by parsing it and move car records accordingly from R001' file into various virtual classification files. Listings of this switch plan are printed in various remote printers for workers to carry out the work order.

After several switch plans have been generated the virtual classification track files will be full of car records. To empty these track files we have to do switch plans

for making up departure trains - to move car records from virtual classification track files to virtual departure track files. Then we can do more switch plans for receiving track files.

When a switch plan is completed by workers and reported back to the yard master, the yard master calls up PLAN and enters the switch plan number that has just completed. PLAN retrieves the switch plan file from the log and this time it moves car records from the real track file R001 to designated real classification track files.

PLAN performs its work and planning on the virtual files to provide the yard master visible results on monitor up to 16 switch plans ahead. Experimentation in the Zhengzhou yard shows users only have to use 5 switch plans ahead to meet their workload productivity target. PLAN uses duplicated sets of track files, switch plan history log files, undo (pre-empt) and redo methodologies to handle roll forward, roll backward predictions and commits. Because of possible unforeseen errors PLAN was implemented with a data synchronization feature to assure data integrity and concurrency. PLAN synchronizes the database by copying all real track files to their corresponding virtual track files and then redo any unexecuted switch plans from the history log in ascending order.

Disk I/O is a problem if a computer generated switch plan must be modified frequently because of a lot of undo and redo of virtual files. For example, if switch plan 1 is found out to be required modified while we are working on switch plan 6, PLAN has to undo six plans (from plan 6, 5, 4, 3, 2, 1 in sequential order), let user modifies switch plan 1 and then redo switch plan 1 to 6 in sequential order. The disk I/O performance can be enhanced by using memory resident car files and indexing

technique to move as little data as possible. Also, the fixed rules and the heuristic rules can always be fine-tuned to provide a perfect switch plan that normally does not require modification.

The "PLAN" expert system was designed to improve the overall productivity of the yard; while relieving the shortage problem of experienced yard masters in Chinese Railway. It also serves as a helpful decision tool even for experienced yard masters. The successful implementation of "PLAN" in the Zhengzhou yard shows that expert system can be applied and used in many railway transportation fields.

Acknowledgement :

We have to thank CIDA and CR for their support in the project. Many thanks to the programming team of the CR's information system, North Jiao Tung University and the Zhengzhou yard for their meticulous and tireless efforts in making the project successful.

References :

- [1] Albert Luk, "A Chinese Computer System For Railway Yard Operation" in Proceedings 1988 International Conference on Computer processing of Chinese and Oriental Languages, 1988 and presented at the conference in Toronto, Canada, Aug. 29 - Sep. 1, 1988.

AUTHOR INDEX AUTEUR

- A -

Ali, M.O., 118
 Attisha, M.R., 103
 Audette, M., 127, 135

- B -

Babineau, D., 60
 Barakat, M., 69
 Becker, A., 36
 Bedetti, A., 139
 Berglund, B.C., 81
 Berkowitz, S., 135
 Bhargava, V.K., 131
 Blevis, B.C., 6
 Bowman, C., 18
 Brokof, U., 10

- C -

Crabb, C.A., 103
 Currie, B., 74

- D -

Daverio, R., 139
 Dawe, B.R., 89
 Delecki, Z.A., 69
 DeSantis, R.M., 108
 DiStefano, B., 139
 DiStefano, E., 139
 Dravnieks, O., 131

- E -

El-Gamal, M.Z., 6, 18

- F -

Feather, T.E., 135
 Fong, C.V., 99

- G -

Gautier, C., 127
 Gomi, T., 143
 Griffith, A., 143
 Guan, L., 135

- H -

Hamburger, D., 74
 Haykin, S., 74, 77
 Hicks, T.A., 53
 Hodgins, D.O., 81
 Houston, S., 32
 Hurteau, R., 108
 Huculak, M., 127

- J -

Janes, R., 29, 60
 Jordan, J.E., 123

- K -

Kelly, M., 89
 Kelly, R.J., 22
 Kerr, R.W., 112
 Khan, R.H., 85
 Knabe, F., 10
 Krasnor, C., 74
 Kroker, H., 69

- L -

Labrèche, L., 57
 Lee-Kwen, A., 135
 Leung, V.C.M., 118
 Little, W.D., 131
 Luk, A., 147

- M -

MacNeil, M., 29
 Manning, P.E., 45
 Mayer, K., 143
 McLane, P.J., 112
 Michaels, J.F., 49
 Mitchell, D.K., 85
 Murfin, A.J., 57

- N -

Nilakantan, A., 127
 Nisson, E., 40
 Nohara, T., 74

- O -

Oppacher, U., 143

- P -

Pawlak, M., 69
 Pesciotta, E.J., 45
 Pozniak, A., 147

- R -

Radhakant, S., 6
 Ram, G.R., 103
 Rudback, N.E., 2
 Ruggiero, J.W., 45

- S -

Sabounghi, L., 143
 Shafai, L., 69
 Singh, K.R., 63
 Smith, I.G., 103
 Smith, A.E., 14
 Soiferman, L., 69
 Spolsky, A.I., 118

- U -

Ulan, R., 6

- V -

Versailles, C.A., 89
 Vickers, M., 32
 Völckers, U., 94